Structure-function studies on Clostridium botulinum neurotoxins

Davies, Jonathan

Award date:
2019

Awarding institution:
University of Bath

Structure-function studies on *Clostridium botulinum* neurotoxins

Jonathan Davies

A thesis submitted for the degree of Doctor of Philosophy
University of Bath, Department of Biology and Biochemistry
September 2018

Declaration of any previous submission of the work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

Much of sections 1.1 and 1.2 have been duplicated verbatim from a published review, Davies *et al.* (2018b).

Declaration of authorship

I am the author of this thesis, and the work described therein was carried out by myself personally.

In memory of Kevin Davies

Acknowledgements

Firstly, I would like to thank my supervisors Prof. Ravi Acharya and Dr. Sai Man Liu their advice, support, and guidance throughout. I also thank Dr. Gavin Hackett for his supervision during the first half of my PhD. This thesis would not have been possible without the joint funding provided by the University of Bath and Ipsen Bioinnovation Ltd. and I am very grateful.

To my colleagues and friends who made this PhD fly by, I thank you all very much and I will certainly miss my time in Bath.

I also send thanks to everyone else who has supported me along the way. To my Mother in particular, I would not be here without her belief and encouragement — I hope you realise just how much I appreciate everything.

Finally, I wish to express my deepest gratitude to Mareike without whom I would have struggled so much, in so many ways — Thank you for everything.

Abstract

Botulinum neurotoxins (BoNTs) are the deadliest proteins known and cause flacid paralysis through inhibition of acetylcholine release from motor neurons. Each BoNT consists of three domains: the binding domain, the translocation domain and the catalytic light chain. The binding domain is responsible for the targetting of neuronal cell membranes and is able to form a dual-receptor complex with cell-surface receptors. The BoNT is taken into the cell through endocytosis where the translocation domain acts to release the catalytic light chain into the cell cytosol. The light chain is then free to cleave one of the SNARE (soluble N-ethylmaleimide-sensitive-fusion-protein attachment protein receptor) proteins, inhibiting exocytosis and thus neurotransmitter release. BoNTs are categorised into serotypes (BoNT/A to /G) and subtypes (e.g. BoNT/A1) based on amino acid variations. BoNTs have many therapeutic uses including as a treatment for cervical dytonia, strabismus, and overactive bladder and the most common commercially available product consists of BoNT/A1. Subtle amino acid variations between BoNT subtypes may have a dramatic impact on the properties of the protein, some of which may be useful for the production of engineered BoNTs in the future.

In this thesis, the crystal structures of the binding domains from BoNT/A1 to /A6 were determined to high resolution and their binding to the protein-receptor SV2 was investigated. The structures of full-length BoNTs in solution were also investigated to confirm whether multiple domain-organisations exist across BoNT serotypes. There are over 150 different BoNTs known and this work also includes a new open-access tool for identifying and retrieving information about specific BoNTs in a fast and easy way.

Contents

List of Figures

List of Tables

Abbreviations

A_{280}	Absorbance at 280 nm
BoNT	Botulinum neurotoxin
BoNT/X	Botulinum neurotoxin serotype X
bont/X	Gene encoding for a botulinum neurotoxin of serotype X
bont	Gene encoding for a botulinum neurotoxin
C. botulinum	*Clostridium botulinum*
DLS	Diamond Light Source
D_{max}	Maximum particle dimension
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
GST	Glutathione S-transferase
H_C	Botulinum neurotoxin binding domain
H_C/X	Botulinum neurotoxin subtype X binding domain

H_N	Botulinum neurotoxin translocation domain
HIC	Hydrophobic interaction chromatography
IEC	Ion-exchange chromatography
IMAC	Immobilized metal affinity chromatography
LB	Lysogeny Broth
LC	Botulinum neurotoxin light chain (protease domain)
LH_N	Botulinum neurotoxin translocation domain and light chain
MAD	Multiple-wavelength anomalous dispersion
MD	Molecular dynamics
MIB	Buffer System: Malonic Acid, Imidazole and Boric Acid (2:3:3 Molar Ratio), pH adjusted with NaOH
MME	Monomethyl ether
mTB	modified Terrific Broth
NSF	N-ethylmaleimide-sensitive fusion protein
PBS	Phosphate buffered saline
PEG	Polyethylene glycol
Rg	Radius of gyration
SAD	Single-wavelength anomalous dispersion
SAXS	Small-angle X-ray Scattering
SNARE	Soluble NSF-attachment protein receptors
SOC Media	Super Optimal broth with Catabolite repression media
SV2	Synaptic vesicle glycoprotein 2
Syt	synaptotagmin
Trx	Thioredoxin
TrxR	Thioredoxin Reductase

CHAPTER 1

Introduction

Much of sections 1.1 and 1.2 have been duplicated verbatim from a published review, Davies *et al.* (2018b). The review is included in full on page 197.

1.1 Bacterial Toxins

Many infectious diseases are caused by proteins secreted by bacterial species. For many bacteria, the secretion of toxins results in a clear benefit to the microbe but for others the benefit remains unclear.

Some bacterial toxins, for which the advantage to the bacteria is unclear, are capable of inducing paralysis that can lead to host death. The tetanus neurotoxin (TeNT) causes such a condition, tetanus, it is produced by *Clostridium tetani*. The anaerobic bacterium, *C. tetani*, is found globally and forms spores which are found commonly in soil. One can speculate as to whether death of the host is beneficial to spore forming anaerobes but the advantage remains a paradox. Another deadly condition, similar to tetanus, and caused by a bacterial toxin is botulism which is characterised by muscle weakness which spreads throughout the body and eventually leads to death (Sobel, 2005). Botulism was first attributed to a toxin by the German doctor, Justinus Kerner, many years before microbes were discovered (Kerner, 1817). Instances of botulism can be categorised into one of three types: food-borne, wound or intestinal. The name botulism originates from the Latin word for sausage, *botulus*. Foodbourne botulism follows the ingestion of contaminated foods such as tinned meals which have not been sufficiently heat sterilised. Wound botulism is caused by infection of open wounds by BoNT-producing bacteria (Fillo *et al.*, 2015). Intestinal botulism, which mainly affects infants, is caused by colonisation of the gut with BoNT-producing bacteria (Umeda *et al.*, 2009). Individuals identified as suffering from botulism must be treated immediately to stop the condition worsening. Treatment consists of mainly supportive measures and in more advanced cases the patient must be intubated and subjected to mechanical ventilation to avoid respiratory failure from the spread of muscle paralysis (Chalk *et al.*, 2014). Anti-toxin derived from equine-sera has been produced which stops further intoxication from further BoNT binding to receptors (Patel *et al.*, 2014). While an extremely rare problem, the condition still affects many people across the world.

1.2 Botulinum Neurotoxins

Botulinum neurotoxins (BoNTs) are produced mainly by *Clostridium botulinum*, under anaerobic conditions (Hatheway, 1990), and are the causative agent of botulism, a rare disease that can lead to paralysis and eventually death. The *C. botulinum* taxon can be divided into four groups (I, II, III and IV), based on phenotypic differences between the bacteria (Collins and East, 1998). *C. botulinum* group I (proteolytic) and group II (non-proteolytic) are mostly responsible for human botulism whereas *C. botulinum* group III is responsible for botulism in other animal species, and *C. botulinum* group IV does not appear to cause botulism (Collins and East, 1998; Carter and Peck, 2015). Across these phenotypes, a range of serologically distinct BoNTs have been identified, based on their recognition by anti-sera, and classified within different serotypes. Up until recently, all BoNTs have been categorised into one of seven serotypes ranging from BoNT/A to BoNT/G. The recent identification of novel BoNTs and BoNT-like proteins, which are not neutralisable by existing anti-sera, has lead to new serotype classifications that do not currently continue from the classical nomenclature *e.g.* BoNT/X. Some *C. botulinum* strains have also been identified which express more than one serotype and some have been found to produce chimeric neurotoxins which may have arisen from recombination events (eg. BoNT/CD and BoNT/DC).

Each BoNT is expressed as a single polypeptide chain of ~ 150 kDa in size (Figure 1.1a) which is cleaved post-translationally by a protease to yield an active di-chain molecule consisting of a ~ 50 kDa light chain (LC) and a ~ 100 kDa heavy chain (HC) linked by a disulfide bond. Some BoNT serotypes are cleaved into a di-chain by an endogenous host protease while others may be cleaved in the target organism (DasGupta and Sugiyama, 1972; Prabakaran *et al.*, 2001). For example, BoNT/A purified from *C. botulinum* culture after 8 h is mostly as a single chain peptide but when purified from a 96 h culture it is in the nicked di-chain form (Dekleva and DasGupta, 1989). The *Clostridial* protease responsible for cleaving BoNT/A has been partially characterised but not yet identified (Dekleva and Dasgupta, 1990). Indeed, for all other BoNTs, no specific protease has been identified that selectively activates the protein. This di-chain molecule is the active form, without proteolytic cleavage into a di-chain the BoNT is much less active. Therefore the disulfide bond connecting the LC and HC is vital for the mechanism of intoxication (Fischer and Montal, 2007; Pirazzini *et al.*,

2011). The LC possesses a zinc-endopeptidase from the M27 family of peptidases, while the HC consists of two domains – a translocation domain (H_N) and a receptor binding domain (H_C). The H_C domain can be further divided into two different folds, a carboxyl-terminal β-trefoil (H_{CC}) with an amino-terminal lectin-like jelly roll containing fourteen β-strands (H_{CN}). The crystal structure of BoNT/A1 shows the domains are arranged in a 'butterfly' arrangement where the LC and H_C are the 'wings' attached to the central H_N 'body' (Figure 1b) (Lacy *et al.*, 1998). Each domain appears to be structurally and mechanistically distinct from one another with the exception of a large loop, termed the belt, which wraps around the LC. The structure of BoNT/B follows the same arrangement of BoNT/A whereas the structure of BoNT/E appears to adopt a 'closed wing' compact conformation where the H_C is rotated around the H_N and LC (Figure 1C) (Swaminathan and Eswaramoorthy, 2000; Kumaran *et al.*, 2009).

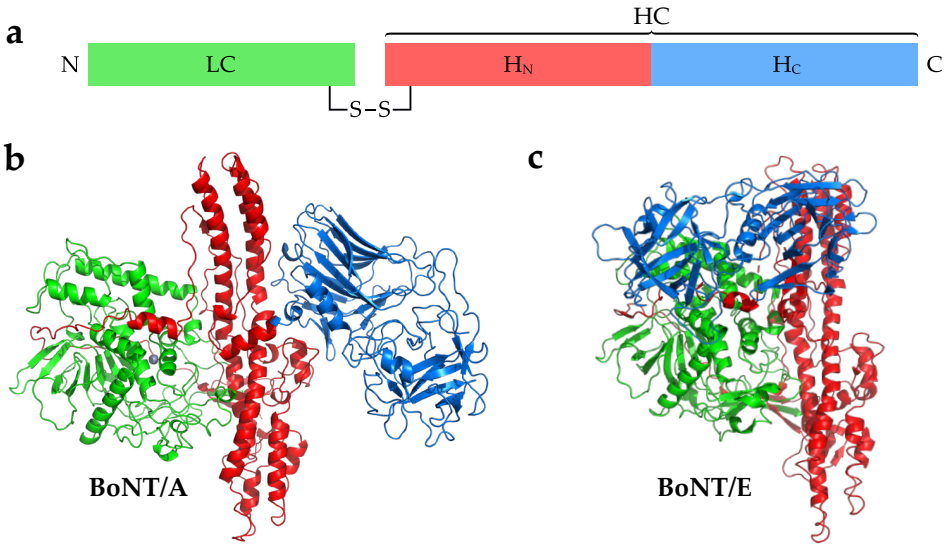


Figure 1.1: Overall BoNT Structure

(a) Representation of general domain layout in BoNTs which includes the protease domain (LC), translocation domain (H_N), and receptor-binding domain (H_C). The LC is linked to the H_N by a single disulfide bond. Full-length crystal structures of BoNT/A (PDB ID: 3BTA) **(b)** and BoNT/E (PDB ID: 3FFZ) **(c)** which exhibit open and closed domain organisation, respectively (Lacy *et al.*, 1998; Swaminathan and Eswaramoorthy, 2000). Figure reproduced from Davies *et al.* (2018b).

1.2.1 The Receptor Binding Domain

The H_C domain is responsible for targeting the protein to the neuronal membrane by binding to receptors present on the neuronal cell surface (Figure 1.2). All classical BoNT serotypes (/A to /G) bind to one or more polysialogangliosides and most also bind to a protein-receptor which together form a dual-receptor complex (Montecucco, 1986; Brunger and Rummel, 2009). The first evidence of ganglioside binding was reported after inhibition of botulism through competition with free ganglioside (Simpson and Rapport, 1971). Polysialogangliosides consist of a hydrophilic complex polysaccharide with many sialic acid residues, bound to a hydrophobic ceramide tail. Different forms of these gangliosides can be found embedded in the cell membrane with the various sugar moieties displayed on the cell surface. The most common examples found on neuronal membranes include GT1b, GD1a, GD1b and GM1. Two types of BoNT protein receptors have also been identified to date: three isoforms of synaptic vesicle glycoprotein 2 (SV2A-C) and two isoforms of synaptotagmin (SytI-II). Both types are involved with the regulated secretion of neurotransmitter from synaptic vesicles (Morgans *et al.*, 2009; Chapman, 2002). SV2A-C contribute to the modulation of exocytosis although their exact role is yet to be determined while SytI and SytII are calcium-sensitive membrane proteins also involved in exocytosis (Morgans *et al.*, 2009; Tang *et al.*, 2006; Chen *et al.*, 2017; Südhof, 2002). Their involvement in synaptic vesicle endocytosis also requires them to be recycled back into the cell through enodcytosis making them excellent targets for BoNTs.

BoNT/A

Within the BoNT/A serotype there are currently eight identified subtypes BoNT/A1 to /A8 which differ by between 3% and 16% at the amino acid level. The most thoroughly characterised BoNT subtype is BoNT/A1 - this is in part due to its use as a therapeutic for several conditions such as spasticity, dystonias, and glabellar lines (Johnson, 1999; Montecucco and Molgó, 2005). The carboxyl-terminal half of the BoNT H_C domain (H_{CC}) contains the peptide motif, H...SxWY...G, which constitutes the core of the ganglioside-binding site (GBS) (Foster, 2014). In contrast to the dual ganglioside binding sites identified on the related TeNT H_C, the GBS of BoNT/A1 can only bind one ganglioside at a time (Rummel *et al.*, 2004b), but is capable of recognising more

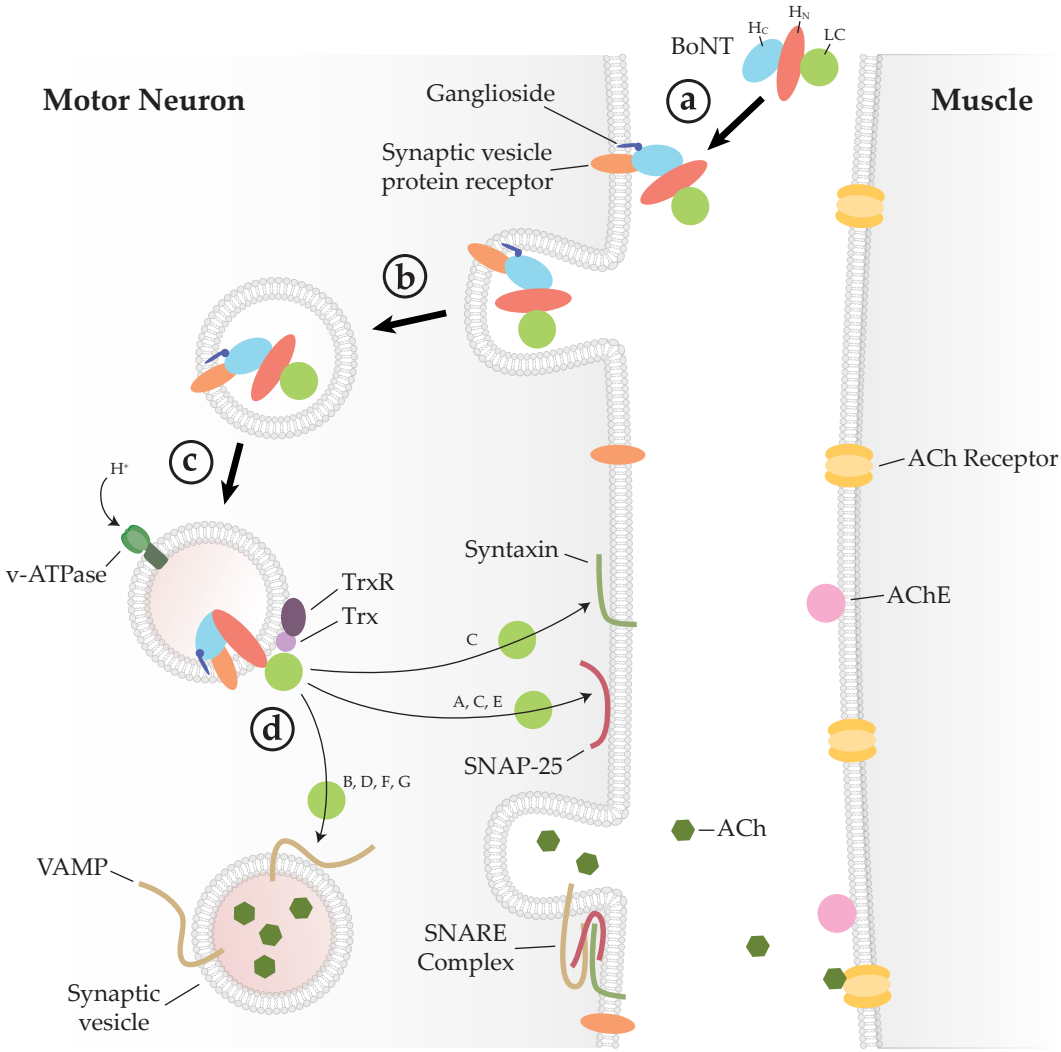


Figure 1.2: Mechanism of BoNT intoxication.

Free BoNT binds the neuronal cell membrane through a dual-receptor complex **(a)**. The BoNT-receptor complex is endocytosed and enclosed within a vesicle **(b)**. A pH gradient within the vesicle causes a conformational change, allowing the LC to be translocated across the membrane **(c)**. Trx, bound to the vesicle membrane, catalyses the reduction of a disulfide bond, allowing the LC to release and cleave its protein target **(d)**. Cleavage of any one of the SNARE proteins disrupts SNARE-mediated exocytosis, stopping ACh release and thus stopping muscle contraction. Labels: TrxR – Thioredoxin reductase; Trx – Thioredoxin; ACh – Acetylcholine; AChE – Acetylcholine esterase; Figure reproduced from Davies *et al.* (2018b).

than one type, specifically GT1b, GD1a, and to a lesser extent GM1 (Yowler *et al.*, 2002; Hamark *et al.*, 2017). The exact interactions between ganglioside and the GBS were first determined from the crystal structure of the H_C domain in complex with GT1b (Stenmark *et al.*, 2008) (PDB ID: 2VU9). This revealed extensive hydrogen–bonding with four of the seven individual monosaccharides within GT1b (Figure 1.3).

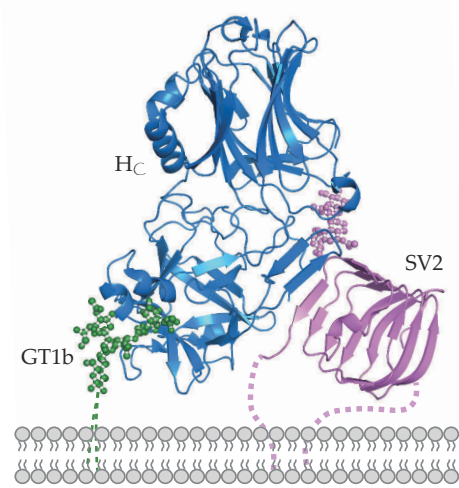


Figure 1.3: H_C/A1 Dual-Receptor Complex
A schematic of the H_C/A1 dual-receptor complex produced by the superposition of two H_C/A1 crystal structure, PDB ID: 2VU9 and 5JLV. The ganglioside, GT1b is shown in green and the protein-receptor, SV2 shown in purple. The N-linked glycan of SV2 is shown as a purple ball and stick representation.

Depletion of gangliosides in neuroblastoma cells completely prevents entry of BoNT/A1 (Yowler *et al.*, 2002), however, gangliosides alone do not mediate cellular entry; for this, BoNT/A also requires the protein-receptor SV2 and in doing so, forms a dual-receptor complex (Dong *et al.*, 2006; Mahrhold *et al.*, 2006) (Figure 1.3). Of the three SV2 isoforms found in humans, BoNT/A1 has the greatest affinity for SV2C (Dong *et al.*, 2006). BoNT/A1 binds specifically to the luminal domain 4 of SV2 (SV2-LD4) via direct backbone-backbone interactions between a β-strand of SV2-LD4 and a β-strand of BoNT H_C (Strotmeier *et al.*, 2014; Benoit *et al.*, 2014), and also through interactions with an N559-linked glycan (Mahrhold *et al.*, 2016). The significance of the latter is highlighted by the inability of BoNT/A1 to bind to bacterially-expressed (*i.e.* non-glycosylated) SV2A or SV2B, and a reduced affinity for non-glycosylated SV2C (Mahrhold *et al.*, 2006; Dong *et al.*, 2008). The crystal structure of BoNT/A1–SV2C-LD4 revealed a large range of interactions, between the H_C and the SV2 glycan, which extended away from the backbone-backbone interactions almost doubling the contact surface area (Yao *et al.*, 2016). There is also some evidence that the H_N domain of BoNT/A1 is also capable of binding to neuronal membrances, although little is known

about the mechanism of translocation and the role of H_N binding (Ayyar *et al.*, 2015).

The other subtypes of BoNT/A (/A1–/A8) are predicted to bind the same receptors as BoNT/A1 due to their high sequence identity between the binding domains. The crystal structure of the BoNT/A2 H_C domain in complex with a non-glycosylated SV2C-LD4 showed that the binding mode is conserved. There are some differences in amino acid sequence between BoNT/A1 and /A2 but SV2C recognition is still conserved as the interaction is mostly mediated by the peptide backbone which is able to tolerate some side-chain variation (Benoit *et al.*, 2017; Gustafsson *et al.*, 2018). There are also some interactions at the SV2 interface which involve protein side-chains. F563 of the SV2-LD4 forms a π-stacking interaction with R1156 of BoNT/A1 while a glutamic acid residue in BoNT/A2 (E1156) causes F563 to adopt a different conformation and the BoNT interacts instead directly with H564 of SV2C-LD4. Such mutations indicate flexibility with respect to the backbone-backbone interaction of SV2 which may also be tolerated due to the extra interactions that occur with the N-linked glycan (Yao *et al.*, 2016). BoNT/A2 has also been shown to have a higher affinity for gangliosides than BoNT/A1 although the interactions mediating this difference have not yet been identified (Kroken *et al.*, 2017). The crystal structures of the BoNT/A3 and /A4 H_C domains suggest a similar mode of interaction to ganglioside compared to BoNT/A1 – the GBS of the former shows a potential loss of a hydrogen bond to one of the terminal sialic acids due to a difference in amino acid (phenyalanine rather than a tyrosine), whereas the latter is conformationally conserved (Davies *et al.*, 2018c). With regard to the SV2 binding site, both structures show slight differences in conformation compared to that of BoNT/A1 due to differences in the primary sequence. Whether this will affect interactions with the SV2 glycan is yet to be determined. Despite high sequence identity between BoNT/A subtypes, significant differences in their intoxication properties have been identified. For example, BoNT/A2 is more potent in neuronal cells than BoNT/A1 possibly due to faster cell entry (Pier *et al.*, 2011; Torii *et al.*, 2011; Pellett *et al.*, 2015) and BoNT/A4 has been reported to be three orders of magnitude less potent than BoNT/A1 (Pellett *et al.*, 2015). It is difficult to attribute these differences to just interactions between the H_C and receptors, and is instead likely as a result of contributions from the H_C, H_N and LC combined.

BoNT/B

There are currently eight subtypes within the BoNT/B serotype (BoNT/B1 to /B8) and they differ by between 1.5% and 7% at the amino acid level (Wangroongsarb *et al.*, 2014). Although the crystal structure of BoNT/B1 exists in an open conformation similar to that of BoNT/A (Swaminathan and Eswaramoorthy, 2000) (Figure 1.1), the BoNT/B serotype targets a different protein receptor on the neuronal cell membrane, namely SytI or SytII (Nishiki *et al.*, 1996; Dong *et al.*, 2003; Rummel *et al.*, 2007; Dong *et al.*, 2007). Crystal structures of the H_C domain from BoNT/B in complex with murine SytII revealed the high specificity of the binding interface where the SytII peptide forms a helix and binds to a hydrophobic groove via six hydrophobic residues (Chai *et al.*, 2006; Jin *et al.*, 2006). Interestingly, BoNT/B displays a much lower affinity toward human SytII than murine SytII due to a single mutation at residue 54 - Phe in rodents and Leu in humans (Strotmeier *et al.*, 2012; Peng *et al.*, 2012; Chai *et al.*, 2006). Considering that SytII is more abundant on human motor neurons than SytI, a significantly larger dose of BoNT/B needs to be administered in order to achieve a similar therapeutic effect as that for BoNT/A. A commercially available formulation of BoNT/B is available which has had lower success, potentially due to the need for increase dosing (Fonfria *et al.*, 2018). To overcome this issue, the BoNT/B binding domain has recently been engineered to increase its binding affinity. Two mutations were identified (E1191M, S1199Y) which showed an 11-fold higher functional efficacy in human cells compared to wild-type BoNT/B1 (Tao *et al.*, 2017). BoNT/B is only capable of entering cells once it has bound to both the Syt receptor and its ganglioside receptor (Figure 1.4), either GT1b or GD1a (Kohda *et al.*, 2007; Rummel *et al.*, 2004b; Nishiki *et al.*, 1996).

The crystal structure of the BoNT/B1 binding domain in complex with both SytII and GD1a show strong interactions with the Sia5 moiety in much the same manner as BoNT/A1 (Berntsson *et al.*, 2013). Although there is no direct contact between SytII and GD1a, there is some evidence that each can influence binding to the other possibly due the spatial arrangement of both binding sites (Atassi *et al.*, 2014). The dual-receptors are what enable the BoNT to target neuronal cell membranes so specifically but there may be another method by which BoNT/B and some other serotypes interact with neuronal membranes. In addition to the dual-receptors, BoNT/B has been reported

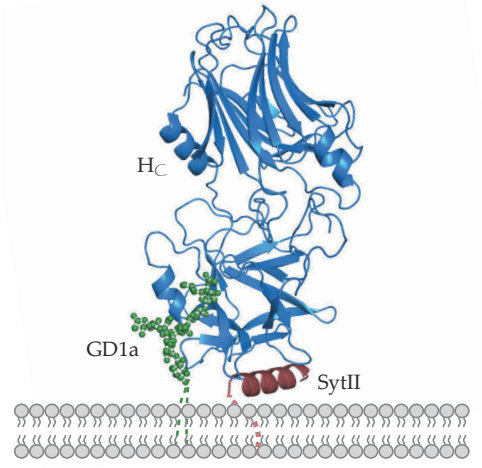


Figure 1.4: H_C/B1 Dual-Receptor Complex
A schematic of the H_C/B1 dual-receptor complex from its crystal structure (PDB ID: 4KBB). The ganglioside, GD1a is shown in green and the protein-receptor, SytII is shown in red.

to interact directly with the cell membrane through an exposed hydrophobic loop (termed the 'lipid-binding loop') located between the ganglioside and Syt binding sites on the H_C (Stern *et al.*, 2018). This interaction may help the BoNT anchor itself to the membrane, increasing the affinity to neuronal cells.

BoNT/C

BoNT/C (specifically known as BoNT/C1) is predominantly associated with botulism in animals rather than humans (Collins and East, 1998; Lindström *et al.*, 2004). There are no other subtypes identified within the BoNT/C serotype – only two distinct protein sequences that have been identified to date (UniProtKB: P18640 (Hauser *et al.*, 1990), Q93HT3 (Takeda *et al.*, 2005)) which both share 99.9% identity. Perhaps confusingly, there are two other botulinum toxins called 'C2 toxin' and 'C3' which are not 'traditional' neurotoxins, but rather refer to different gene products – a binary AB toxin and an exoenzyme, respectively (Stiles *et al.*, 2014; Chellapandi and Prisilla, 2017; Evans *et al.*, 2004). The mechanisms of cell-binding is of great interest despite the association with non-human botulism because unlike the majority other BoNTs, no protein receptor for BoNT/C1 has yet been identified (Tsukamoto *et al.*, 2005; Kroken *et al.*, 2011a). Interestingly, while the conserved SxWY ganglioside-binding motif is absent from the H_C domain, BoNT/C1 is still able to bind gangliosides (Rummel *et al.*, 2009). Indeed, an extended hydrophobic loop termed the 'ganglioside binding loop' (GBL) was reported to be essential for neuronal binding, however, the specific

interactions to a ganglioside have yet to be determined (Kroken *et al.*, 2011a). Crystal structures of the BoNT/C1 H_C domain in complex with sialic acid – a monosaccharide which may indicate a the potential for ganglioside binding – revealed two potential binding sites that are spatially independent of the GBS identified in other BoNTs (Strotmeier *et al.*, 2011; Karalewitz *et al.*, 2012).

BoNT/D

Like BoNT/C, there are no subtypes of BoNT/D, of which there are multiple sequences that share a primary sequence identity of $> 96\%$. BoNT/D appears to recognise all three isoforms of SV2 (Rummel *et al.*, 2009; Strotmeier *et al.*, 2011). Cells lacking SV2 do not get intoxicated by BoNT/D, however this can be restored by the expression of any of the three SV2 isoforms (SV2A, B, C) (Peng *et al.*, 2011). It was further demonstrated that SV2A/B knockout neurons displaying a chimeric form of SV2-LD4 (SV2A, B or C) alone were unable to mediate BoNT/D entry despite rescuing intoxication for other serotypes including BoNT/A and /E. Mutation of the N537 N-linked glycosylation site of SV2 also had no effect on BoNT/D entry despite blocking entry to BoNT/E (Peng *et al.*, 2011). This suggests that the SV2 receptor binding domain in BoNT/D may be distinct from other SV2-interacting BoNTs such as BoNT/A and the glycan interaction plays less of a role. Gangliosides are also required for BoNT/D cell entry (Strotmeier *et al.*, 2010), however like BoNT/C, BoNT/D does not contain an SxWY motif in the GBS although the site is still able to recognise gangliosides (Zhang *et al.*, 2010). It is also proposed to contain a second binding site termed Sia-1, since mutation of this site results in reduced ganglioside binding (Kroken *et al.*, 2011b).

BoNT/E

There are currently twelve known BoNT/E subtypes (BoNT/E1 to /E12) whose amino acid identities vary by up to 12%. The protein receptor for BoNT/E is SV2, although only isoforms SV2B and SV2C are capable of mediating entry (Dong *et al.*, 2008; Stefan *et al.*, 2013), and in the presence of gangliosides (Kamata *et al.*, 1986). The SxWY motif is conserved in the BoNT/E GBS and direct binding of GT1b has been observed (Sun *et al.*, 2012). No crystal structures of BoNT/E in complex with receptor or ganglioside

have yet been solved, therefore, the precise molecular basis of their interactions have yet to be determined. The native crystal structure of BoNT/E has been solved and it reveals a conformation that is significantly different to that of BoNT/A and BoNT/B (Kumaran *et al.*, 2009). As previously discussed, the H_C domain wraps around the toxin giving the protein more compact shape overall (Figure 1.1). BoNT/E is capable of entering cells much faster than BoNT/A (Keller *et al.*, 2004), and this domain organisation has been proposed to prime the toxin for translocation resulting in a faster onset of paralysis (Kumaran *et al.*, 2009). However, investigations using various chimeras of BoNT/A1 and BoNT/E1 showed that the speed of translocation is not affected by the H_C domain (Wang *et al.*, 2008).

BoNT/F

In addition to BoNT/F1, there are eight other BoNT/F subtypes (BoNT/F2 to /F9) which differ by up to 30% sequence identity. The exact protein receptor for BoNT/F1 has been reported to be glycosylated SV2 (Fu *et al.*, 2009; Rummel *et al.*, 2009) but this remains to be established conclusively. For example, one study showed that BoNT/F activity decreased when H_C/A was introduced as a competitor molecule (Rummel *et al.*, 2009), whereas a separate study demonstrated that BoNT/F1 entry in neurons was unaffected by a double SV2A/B knockout in cortical neurons (which have negligible expression of SV2C) (Yeh *et al.*, 2010). For ganglioside binding, BoNT/F1 requires gangliosides containing an α2,3-linked sialic acid on the terminal galactose, *i.e.* GT1b or GD1a (Fu *et al.*, 2009; Rummel *et al.*, 2009). The SxWY motif is conserved in BoNT/F and the crystal structure of the H_C domain from BoNT/F1 in complex with GD1a confirmed the existence of a GBS (Benson *et al.*, 2011). The evidence for a protein receptor of BoNT/F is inconclusive and further insights into neuronal binding are needed, possibly through a structural approach.

BoNT/G

Only two protein sequences of BoNT/G are currently known to exist, and they share 99.9% amino acid identity. The protein receptor for BoNT/G is either SytI or SytII although interestingly the interface diverges from BoNT/B and it has been found to

possess a lower binding affinity (Dong *et al.*, 2007; Stenmark *et al.*, 2010; Willjes *et al.*, 2013). Only 5 of 14 residues involved in the BoNT/B–SytII interaction are conserved in BoNT/G (Rummel *et al.*, 2007; Stenmark *et al.*, 2010). Like BoNT/B, BoNT/G also displays a low affinity for the human SytII receptor due to a human/chimpanzee specific mutation (Strotmeier *et al.*, 2012). The BoNT/B H_C domain was successfully engineered to improve human SytII binding and a similar approach would be worth investigating here if BoNT/G were to be assessed as a candidate for a human therapeutic (Tao *et al.*, 2017). BoNT/G possesses the conserved SxWY motif in its GBS and binds preferentially to GT1b (Schmitt *et al.*, 2010). In addition to the dual-receptor interactions, BoNT/G also contains a 'lipid-binding loop' (residues 1252-1256) similar to that of BoNT/B which can directly interact with the cell membrane to further contribute binding affinity (Stenmark *et al.*, 2010; Stern *et al.*, 2018), and deletion of this loop has been shown to dramatically decrease neurotoxicity (Stern *et al.*, 2018).

BoNT/CD

BoNTs composed of domains or regions from multiple different serotypes also exist in nature. BoNT/CD is a mosaic toxin composed of a LC domain and a H_N domain that is most similar to BoNT/C, and a H_C domain that is most similar to BoNT/D. Interestingly the binding domain of BoNT/CD binds synaptosomes more tightly than BoNT/D (Tsukamoto *et al.*, 2005). This may be due to a few specific residues, K1118 and K1136 (which differ from the equivalent residues in BoNT/D, E1114 and G1132), since mutation of these lysines results in a dramatic loss in synaptosome binding affinity (Zhang *et al.*, 2011). The role of these residues in binding is unknown. Protein residues which may also interact with a ganglioside have also been identified through crystallisation with a sialic acid molecule (Zhang *et al.*, 2013).

BoNT/DC

The BoNT/DC chimeric protein possesses a LC domain and a H_N domain with 96% sequence identity to BoNT/D, and a H_C domain similar to that of BoNT/C (74% sequence identity) (Moriishi *et al.*, 1996; Nakamura *et al.*, 2010). Botulism caused by BoNT/DC is usually found outside of humans in birds and other mammals but it has

also been found capable of binding human neuronal cells (Nakamura *et al.*, 2010; Peng *et al.*, 2012). Despite having a binding domain similar to BoNT/C, BoNT/DC targets a protein receptor and binds to either SytI or SytII. This interaction is mediated by hydrophobic residues, and is distinct from that of BoNT/B (Berntsson *et al.*, 2013). The BoNT/DC protein is particularly interesting as it appears it may not require complex gangliosides to enter target neurons (Zhang *et al.*, 2017a; Karalewitz *et al.*, 2010). However, the crystal structure of BoNT/DC in complex with Sialyl-T suggests that BoNT/DC is capable of recognising a single sialic acid and thus potentially a range of gangliosides. The structure also reveals the presence of an extended 'lipid-binding loop' that is also observed in BoNT/B and BoNT/G (Zhang *et al.*, 2017a; Stern *et al.*, 2018).

BoNT/FA(HA)

BoNT/FA was recently identified in 2014 from a case of infant botulism (Barash and Arnon, 2014; Dover *et al.*, 2014). At the time it was referred to as BoNT/H (and sometimes still as BoNT/HA) due to its non-neutralisable antigenicity (Barash and Arnon, 2014) and phylogenetic analysis of the *bont* sequences placed the gene in a lineage distinct from other serotypes (Dover *et al.*, 2014). After a protracted period of data restriction due to supposed safety concerns (Hooper and Hirsch, 2014; Relman, 2014; Johnson, 2014) the sequence was finally released to the scientific community after which it was determined that the molecule was a mosaic toxin composed of a LC similar to that of BoNT/F5, an H_N domain similar to that of BoNT/F1, and an H_C domain simliar to that of BoNT/A1 (Gonzalez-Escalona *et al.*, 2014). Direct binding of the BoNT/FA HC domain has been confirmed for glycosylated SV2C-LD4 (Yao *et al.*, 2017) and crystal structures of this binding domain show some slight differences with respect to BoNT/A1 which would be consistent with a decreased affinity towards the protein backbone of SV2 (Yao *et al.*, 2017; Davies *et al.*, 2018a). Although no ganglioside-bound structure of BoNT/FA has yet been solved, the structure of the GBS appears to maintain the same fold as that observed for BoNT/A1 (Davies *et al.*, 2018a). SV2 is likely the protein receptor for BoNT/FA and direct binding has been confirmed for glycosylated SV2C-LD4 (Yao *et al.*, 2017). The BoNT/FA sequence contains mutations with respect to BoNT/A1 which result in decreased affinity towards the protein backbone of SV2, as determined by a pull-down assay against non-glycosylated SV2C,

while the equivalent residues involved in glycan binding remain unchanged (Yao *et al.*, 2017). The effect of these mutations towards different isoforms of SV2 remains to be seen. The ganglioside-binding site is able to maintain the same fold as BoNT/A1 but no ganglioside bound structures yet exist so the exact interactions remain to be determined (Davies *et al.*, 2018a). In recent assays using cultured rat embryonic spinal cord neurons and rat cortical neurons, BoNT/FA was found to be much more potent than BoNT/A1. However, counter intuitively the toxin was much less potent when assayed using an *ex vivo* mouse phrenic nerve hemidiaphragm (mPNHD). These results along with the methods used for each assay point toward a toxin that may have a slow speed of onset despite a highly active LC (Hackett *et al.*, 2018). Understanding interactions of BoNT/FA with its receptors are crucial to both determine what causes intoxication differences and for developing novel therapeutics.

BoNT/X

Recent bioinformatic analysis of online protein databases identifed a novel BoNT serotype termed BoNT/X (Zhang *et al.*, 2017b). *C. botulinum* strain 111, that already was known to express BoNT/B, was found to contain the gene for another BoNT molecule that shared low primary sequence identity to other serotypes (<30%) and was previously unidentified. It is unknown yet whether this molecule is capable of causing human botulism. This suggests that this toxin significantly diverged from other serotypes during its evolution, despite this, recent structural characterisation of the LC has revealed a core fold common to all BoNTs (Masuyer *et al.*, 2018). Little is known about the BoNT/X H_N and H_C domains and attempts are underway to determine the specific receptor/s that it targets and how it functions *in vivo*. The BoNT/X H_C does contain an SxWY sequence motif indicating it potentially shares similar ganglioside binding characteristics with other BoNTs. Due to its divergence and low sequence similarity to existing BoNTs, structural and functional characterisation could lead to new insights into receptor binding which could be exploited for future therapeutics.

BoNT-like proteins

Considering that BoNTs are the deadliest biological agents that exist, it was surprising to find BoNT-like proteins produced by non-*Clostridium* species. The first was found in 2015 and is now referred to as 'BoNT/Wo', named after the bacterium that produced it, *Weissella oryzae* SG25 (Tanizawa *et al.*, 2014; Mansfield *et al.*, 2015). Characterisation of the BoNT-like protein determined its differences to be significant enough that it is not recognised using existing BoNT antisera, specific for BoNT/A – BoNT/G, although some cross-reactivity with anti-BoNT/C and anti-BoNT/D was noted (Zornetta *et al.*, 2016). BoNT/Wo does not contain any typical BoNT motifs in the receptor-binding domain (Zornetta *et al.*, 2016). This is be consistent with zero reported cases of botulism in humans; indeed, it has been speculated that BoNT/Wo may instead target SNARE-mediated plant defence systems (Mansfield *et al.*, 2015). Identification of the receptors for BoNT/Wo may shed new light on alternative receptor-binding mechanisms that BoNTs and BoNT-like proteins can utilise. More recently, another BoNT-like gene cluster was discovered in the bacterium *Enterococcus faecium*, which is a ubiquitous commensal microorganism commonly found in the gut of mammals. The BoNT-like protein, referred to as BoNT/En (or eBoNT/J), possesses many traditional BoNT motifs, including a HExxH zinc-binding motif in the LC and a ganglioside-binding SxWY motif in the HC domain (Zhang *et al.*, 2018; Brunt *et al.*, 2018). Early studies indicate that rodents do not possesses the receptor/s for BoNT/En intoxication (Zhang *et al.*, 2018) and so it is unlikely to also affect humans.

1.2.2 Translocation

The mechanism of BoNT intoxication is a multipart process, after neuronal cell binding the BoNT is internalised while attached to its receptors into an endocytic vesicle. The protease action of the LC is of no use within a vesicle as its target peptides are within the cytosol, thus the BoNT must translocate the LC through the vesicle membrane. This process is the least well understood part of the intoxication process and recent advances in techniques such as cryo-electron microscopy have yet to uncover the exact process. Following endocytosis, the vesicle matures into an endosome and the endosomal pH is decreased by proton pumps within the vesicle membrane to pH $\approx$ 5.4 (Overly *et al.*, 1995), which causes BoNTs to undergo a conformational change.

Structural changes of the entire BoNT have been observed in response to pH change, and it appears that these changes are reversable (Chellappan *et al.*, 2015). The exact mechanism of translocation is still not well understood but it is proposed that the H_N forms a pore through which a partially unfolded LC passes into the neuronal cytosol (Koriazova and Montal, 2003; Pirazzini *et al.*, 2016; Fischer *et al.*, 2012). Evidence of a transmembrane pore, whose internal diameter is estimated to be ∼15 Å, is too small to accommodate the fully-folded LC (Cai *et al.*, 2006). In order to cross the pore, the LC must partially unfold, likely into a molten globule state (Cai *et al.*, 2006), and reassemble into a function protease in the cytosol. The other BoNT domains, the H_C and H_N, may act as chaperones that helps to drive the LC translocation through a pore (Fischer, 2013). Following translocation, the LC remains bound to the H_N on the cytosolic side due to a single disulfide bond. The bond bridges two cysteine residues close to a surface exposed loop which is cleaved post-translation leaving the LC and HC linked only by the disulfide bond (Figure 1.5).

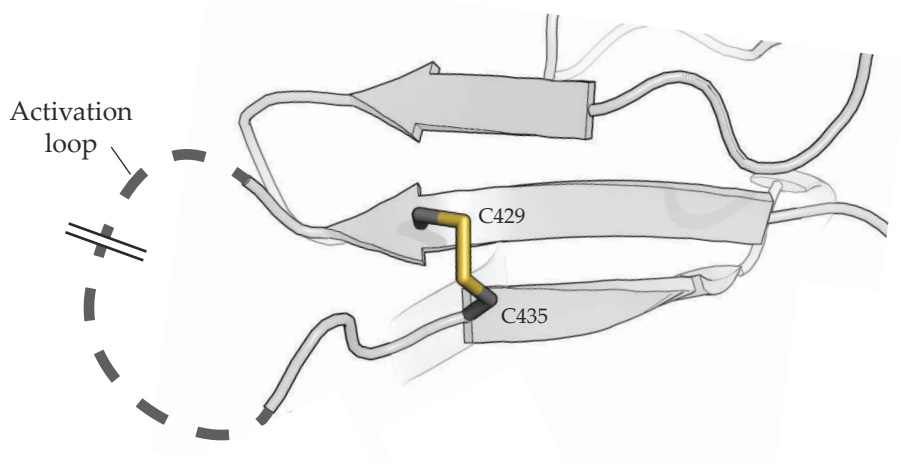


Figure 1.5: BoNT Activation Loop and Inter-Chain Disulfide Bond
The activation loop (dashed line) and C429-C435 disulfide bond from the BoNT/A1 crystal structure (PDB ID: 3BTA) (Lacy *et al.*, 1998). Parallel lines on the activation loop indicate the general area where the peptide is nicked into a di-chain activated protein.

The specific protease responsible for the inter-chain proteolysis is unknown though likely a host protease (Dekleva and Dasgupta, 1990), the chain may also be cleaved by other proteases such as trypsin and Lys-C as is the case when producing recombinant

BoNTs (Hackett *et al.*, 2018). Single-chain BoNTs may still have some residual activity (Shi *et al.*, 2008), but in the case of therapeutic molecules this would be inefficient. The reduction of the interchain disulfide bond is mediated by thioredoxin (Trx), which itself is regenerated by thioredoxin reductase (TrxR). Disulfide cleavage is essential to full intoxication, inhibition of Trx is sufficient to block the LC release (Pirazzini *et al.*, 2014; Zanetti *et al.*, 2015).

1.2.3 Cleavage of SNARE Target

The paralytic action of the BoNT is a result of the LC alone, and the other domains play no role after binding and translocation (de Paiva and Dolly, 1990). The BoNT LC is a zinc-dependent protease which cleaves a soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE). Cleavage of any one of the three SNARE proteins is sufficient to block SNARE complex formation, preventing vesicle-plasma membrane fusion and thus inhibiting exocytosis and release of acetylcholine, causing flaccid paralysis. The structure of the LC is highly conserved across all BoNTs despite some sharing as little as 30% amino acid identity. BoNT/A, /C and /E cleave the protein SNAP-25, while BoNT/B, /D, F, and /G cleave VAMP1, 2, and 3 (Schiavo *et al.*, 1993a; Williamson *et al.*, 1996; Schiavo *et al.*, 1992; Schiavo *et al.*, 1993b; Schiavo *et al.*, 1994). The BoNT/C LC the outlier of all LCs as it is able to additionally cleave Syntaxin I (Blasi *et al.*, 1993b). Each BoNT LC possesses extremely high specificity for its target SNARE protein in all cases, cleaving at a single peptide bond within the SNARE (Rossetto *et al.*, 1994). The first crystal structure of a BoNT LC in complex with a SNARE peptide was determined by Breidenbach and Brünger (2004), and it revealed a range of exosites which are responsible for the remarkable specificity (Figure 1.6).

The recent emergence of novel BoNT serotypes and BoNT-like proteins has increased the repertoire of potential SNARE targets. BoNT/Wo, while not thought to induce botulism in humans, is capable of cleaving human VAMP at a unique location (Trp89–Trp90) (Zornetta *et al.*, 2016). In addition, following the discovery of BoNT/X, its LC was investigated and found to cleave non-canonical substrates such as VAMP4, VAMP5 and Ykt6 while still containing the core fold common to all BoNTs (Zhang *et al.*, 2017b; Masuyer *et al.*, 2018).

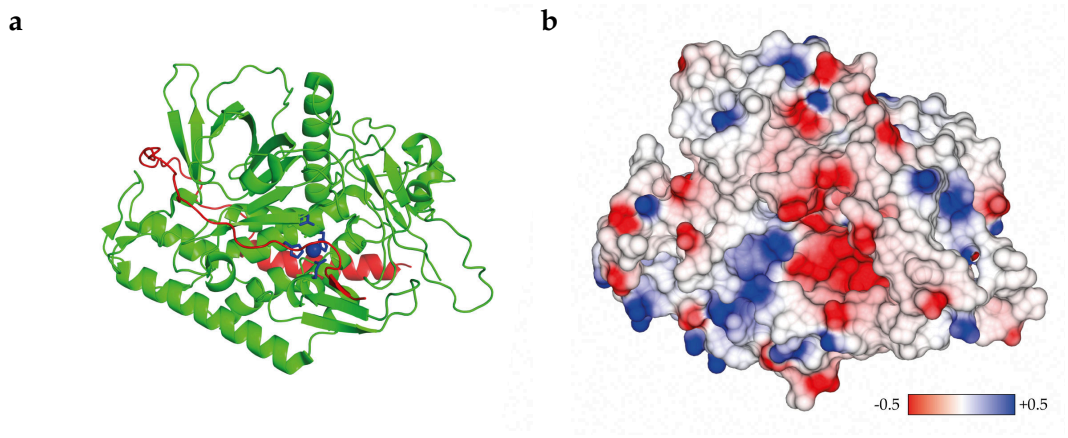


Figure 1.6: LC/A1 in Complex with SNAP-25
(a) Crystal structure of BoNT/A1 LC (green) in complex with its SNARE target, SNAP-25 (red; PDB ID: 1XTG, Breidenbach and Brünger, 2004). **(b)** Surface representation displaying the surface charge of the LC in the absence of the SNAP-25 peptide.

At high doses, some BoNTs can not only cause blockage of synaptic vesicle release but also result in cell death (Peng *et al.*, 2013). The SNARE proteins Syntaxin and SNAP-25 are both essential for membrane fusion functions that are independent of synaptic vesicle release and near complete loss of them can result in neuronal cytotoxicity (Peng *et al.*, 2013; Berliocchi *et al.*, 2005).

1.2.4 Duration of Action

The duration of action between different BoNT serotypes and even high identity subtypes can vary drastically. The cause of this variation is not fully understood and may be influenced by a range of factors. An example of duration variation between subtypes was demonstrated through a long-term study of BoNT/A1 - /A5 activity within a mouse model. It was determined that BoNT/A1, 2, 4, and 5 have a similar duration of action (> 10 months), while the duration of BoNT/A3 is significantly shorter (around 3 months) (Pellett *et al.*, 2015). This observation has also been confirmed in cultured neurons (Whitemarsh *et al.*, 2014). The mechanism controlling the duration of action is not fully understood but it is likely a result of the LC alone. An

engineered hybrid toxin containing the LC/A1 and LH_N/A3 was found to maintain a long duration of action whereas a hybrid containing LC/A3 and LH_N/A1 resulted in a much shorter duration (Pellett *et al.*, 2018). There is some evidence to suggest that the localisation of the LC within the cytosol is a significant factor that determines its duration and efficacy, and as such, mutations that affect localisation have been identified. The LC of BoNT/A1 is known to localise at the plasma membrane within the cytosol while A3 is spread throughout the cytosol.

The terminal residues of LC/A1 may contain signals which direct the LC to the plasma membrane, in particular, disruption of a dileucine motif at the C-terminus (D/EXXXLL) present in LC/A may cause the LC to become more diffuse in the cytosol (Fernández-Salas *et al.*, 2004). Mutation of residues in and around the dileucine motif can be used to modulate the duration of action of BoNT/A1 (Scheps *et al.*, 2017). This alone cannot fully explain the differences between BoNT/A1 and /A3 as both contain the dileucine motif. All BoNT/A subtypes contain the motif whereas it is not present in any other BoNT, as other serotypes also display varying durations with respect to each other, another mechanism may be at play. The N-terminus of the LC has also been implicated with membrane localisation. N-terminal residues of BoNT/A1 LC can interact with high-affinity directly with the SNARE target, SNAP-25, present at the plasma membrane. An N-terminal truncation of 7 amino acids causes the LC to become more diffuse in the cytosol rather than strongly associated with SNAP-25 at the plasma membrane, but intriguingly does not affect the longevity of SNAP-25 cleavage (Chen and Barbieri, 2011). A double N-terminal mutant of BoNT/A1 LC (K6A, K11A) also results in loss of association at the membrane, the mechanism for this is not fully understood (Chen and Barbieri, 2011). Interestingly, BoNT/A3 is the only BoNT/A subtype that has a mutation at K11R. The ability to remain bound at the plasma membrane may help the LC to avoid proteolysis and increase the protein half-life, remaining active for a longer period of time but it is certainly not the only cause.

Another mechanism to increase the half-life and avoid degradation within the cytosol may involve avoiding ubiquitination by host ubiquitin-ligases. A protein implicated in ubiquitination, TRAF2, has been found to associate directly with BoNT/E LC, a short-lasting BoNT, promoting proteasomal-degradation within the cell, while BoNT/A LC appears not susceptible (Tsai *et al.*, 2010). More recent evidence suggests that

the BoNT/A LC is also targetted by a ubiquitin-ligase, HECTD2, but it is able to evade ubiquitin-dependent degradation by direct association with the deubiquitinating enzyme VCIP135/VCPIP1 (Tsai *et al.*, 2017). Together, these data suggest that the localisation of the LC within the cytosol, and its ability to avoid degradation by host proteolysis, are what drives the long duration of action identified in some BoNT subtypes.

Table 1.1: BoNT Receptors and SNARE-Targets

Serotype	Protein-receptor	Ganglioside-recptor	SNARE-target
BoNT/A	SV2A[1], SV2B[1], SV2C[1,2]	GT1b[3,4], GD1a[3], GM1[3]	SNAP-25[5,6]
BoNT/B	SytI[7], SytII[7]	GD1a[8], GT1b[8]	VAMP[9]
BoNT/C	None identified†	GT1b[10,11] = GD1b[10,11] > GD1a[11] > GM1a[11]	SNAP-25[12], Syntaxin[13]
BoNT/D	SV2A[14], SV2B[14], SV2C[14]	GM3[15], GM2[15], GM1[15] and GD1a[15], or GM3[15]	VAMP[6]
BoNT/E	SV2A[16], SV2B[16]	GD1a[17], GT1b[17], and GQ1b[17]	SNAP-25[6]
BoNT/F	SV2A[18,19], SV2B[18,19], SV2C[18,19]	GT1b[18,19] = GD1a[19] > GM3[19]	VAMP[20]
BoNT/G	SytI[21], SytII[21]	Unknown preference[22]‡	VAMP[23]
BoNT/X[24]			VAMP[24], Ykt6[24]

† BoNT/C utilises dual ganglioside binding sites and has no protein receptor.

‡ While gangliosides have been confirmed as a receptor, the ganglioside preference is yet to be determined.

[1] Dong *et al.* (2006), [2] Mahrhold *et al.* (2006), [3] Yowler *et al.* (2002), [4] Rummel *et al.* (2004b), [5] Blasi *et al.* (1993a), [6] Schiavo *et al.* (1993a), [7] Nishiki *et al.* (1994), [8] Ogasawara *et al.* (1991), [9] Schiavo *et al.* (1992), [10] Tsukamoto *et al.* (2005), [11] Karalewitz *et al.* (2012), [12] Williamson *et al.* (1996), [13] Blasi *et al.* (1993b), [14] Peng *et al.* (2011), [15] Strotmeier *et al.* (2010), [16] Dong *et al.* (2008), [17] Kamata *et al.* (1993), [18] Rummel *et al.* (2009), [19] Fu *et al.* (2009), [20] Schiavo *et al.* (1993b), [21] Rummel *et al.* (2004a), [22] Dong *et al.* (2007), [23] Schiavo *et al.* (1994), [24] Zhang *et al.* (2017b)

1.3 Applications of BoNTs

Although BoNTs are deadly toxins with the potential for misuse as bioweapons, they are perhaps surprisingly best known for their use in cosmetic applications (Hexsel *et al.*, 2011). For over 30 years the main application has been the treatment of glabellar lines, however recently the powerful action of BoNTs has been exploited for many therapeutic applications including hyperhidrosis and spasmodic dysphonia (Bhidayasiri and Truong, 2008; Watts *et al.*, 2008; Kessler *et al.*, 1999). There are currently four different BoNT formulations (Dysport, Botox, Myobloc and Xeomin) which are each approved for use in multiple indications within the United States and Europe (Kane *et al.*, 2015; Fonfria *et al.*, 2018).

Currently all approved BoNT formulations include protein that has been purified directly from its associated *C. botulinum* strain and all apart from Xeomin contain addition neurotoxin associated proteins. When considering BoNTs as therapeutic agents, it would be advantageous to produce a cleaner product, free of contaminating proteins, so as to minimise the potential for an immunogenic reaction (Benecke, 2012). Recombinant neurotoxins provide the answer to a simpler production method and expression systems including *E. coli* are being explored. There is currently one recombinant BoNT in phase I clinical trials (Fonfria *et al.*, 2018).

The modular structure of BoNTs allows flexibility with respect to protein engineering and this has been exploited to produce novel therapeutics that are capable of targeting different cell types. By replacing the binding domain with a different targetting peptide the BoNT can be designed to inhibit SNARE mediated exocytosis from cells other than neuronal cells . These novel proteins have been termed Targetted Secretion Inhibitors (TSIs). One example of a TSI is a protein which consists of the LC and H_N domains from BoNT/D and a fragment of the growth hormone releasing hormone (GHRH) which targets pituitary somatotroph cells through binding to the GHRH-receptor (Masuyer *et al.*, 2015). By inhibiting vesicle release from pituitary somatotroph cells this molecule can be used to investigate acromegaly (Garcia *et al.*, 2013)

1.4 Summary

BoNTs are highly specific and potent exotoxins that can be exploited for therapeutic gain. Our knowledge of the molecular aspects of botulinum neurotoxin, such as mechanism of cell targeting and internalisation is incomplete. We have yet to fully understand the binding mechanism of others and also how subtle amino acid differences may result in differences of intoxication (*i.e.*, between subtypes). From what we know so far, X-ray crystallography has suggested that the mechanism of binding is more complex than was initially thought, and many serotypes have evolved unique mechanisms of receptor recognition. It is possible that BoNTs, such those as from the /A, /D and /E serotype may accommodate heterogeneous glycosylation of its protein receptor and all may target a variety of gangliosides to ensure successful binding to the neuronal cell membrane. It is not a trivial task to determine how BoNTs bind to their receptors on neuronal cell membranes, especially when trying to replicate the conditions *in vivo*. With the recent discovery of new BoNTs and BoNT-like molecules in other bacterial species, this raises questions regarding the evolution of the *bont* gene cluster, their ability to be transferred between species, the potential implications to biosafety, and the need for an agreed, consistent naming convention to avoid confusion and ambiguity (Mansfield and Doxey, 2018; Williamson *et al.*, 2017). Fast characterisation of exiting subtypes along with each of these novel toxins would provide a greater understanding of the mechanisms employed for intoxication and how different BoNTs utilise distinct methods. In particular, atomic data from individual subtypes would help to reveal the causes of different intoxication properties and could be used for structural and functional analyses, providing insights for the design of novel BoNTs (Kammerer and Benoit, 2014).

1.5 Aims

1. **Investigate how sequence variation affects the BoNT receptor binding domain structure.** BoNTs are extremely potent neurotoxins that have a tremendous potential for treating a number of conditions. They first target neuronal cell membranes and bind receptors on the cell surface. Very little research has focussed on how the sequence variation between subtypes affects the structure and function of the receptor binding domain. The H_C domains from six BoNT/A subtypes will be produced in order to investigate whether they were suitable for structural studies. Crystallisation will be attempted for each and structures will provide an insight into the affect of subtle differences on receptor binding and become the basis for future modelling of engineered recombinant proteins.

2. **Determine the domain organisation of full-length BoNTs in solution.** Following receptor binding, BoNTs must enter the cytosol of the cell. The translocation of BoNTs into the cytosol after neuronal cell entry is one of the key steps of intoxication for which very little is known. The crystal structure of BoNT/E unveiled a novel conformation of the protein compared to the existing BoNT/A structure. It has been suggested that the 'closed' conformation is what allows BoNT/E to translocate faster into the cytosol than other BoNTs. Investigation of the solution structure will be performed to confirm whether the crystal structure is representative of the BoNT *in vivo*.

3. **Generate a BoNT protein sequence database.** Finally, to further highlight BoNT variation and in an effort to improve reporting of the exact BoNT protein sequence used in research, we aimed to put together an online database.

CHAPTER 2

X-ray Crystallography

2.1 Introduction

X-ray crystallography is a powerful technique that is used to uncover and understand the molecular structure of molecules including proteins. Determining the structure of a protein allows researchers to deduce the function and other aspects of the mechanism of action through careful analysis of atomic positions. Existing knowledge of a protein and its function as determined through biochemical characterisation can complement a protein crystal structure leading to full understanding of the biochemical process that it is involved in.

Proteins are extremely small and generally range between a few and tens of nanometres – too small to view by methods such as light microscopy due to the wavelength of visible light. Electromagnetic radiation in the visible range is useful for characterising protein samples using techniques such as circular dichroism (CD) or dynamic light scattering (DLS) but these methods are not able to produce precise information about the atomic positions of atoms within the molecule. Instead X-rays, with their wavelength being much shorter than visible light, can be used to visualise proteins. X-ray crystallography exploits a property of crystals called diffraction where photons interact with repeating planes of atoms and are reflected where they can be detected. The first crystal structure determined was that of myoglobin in 1958 by John Kendrew *et al.*, and since then over 130,000 protein structures have been released publicly in the Protein Data Bank (PDB). Crystallography, as the name suggests, requires crystals of the molecule of interest to be produced – this is the major challenge of the technique, one which, while now capable of automation, still poses a challenge more than half a century after the first structure. Other techniques for studying protein structure such as NMR or cryo-electron microscopy avoid the need for crystals but have their own drawbacks. In general, NMR has an upper size limit of ~ 40 kDa for the protein under investigation while cryo-electron microscopy has a lower size limit of ~ 150 kDa.

Many of the results presented in this thesis make extensive use of X-ray crystallography and so an overview of the technique and theory is given below.

2.2 Crystals and Symmetry

The production of proteins crystals or other small molecule crystals is usually a challenging task and the exact conditions that result in crystallisation must be experimentally determined. It is now common to screen hundreds, if not thousands, of conditions to find the ideal environment to crystallise a specific protein (Stock *et al.*, 2005). One method to produce protein crystals is by vapour diffusion (Figure 2.1) (Luft *et al.*, 2014). Here, a small volume of concentrated protein solution is mixed with another small volume of the screening condition. A larger reservoir of the screening solution is sealed within the same well as the protein solution. Due to the dilution of the screening condition when mixed with protein, the reservoir will contain a higher concentration of precipitant and vapour diffusion will occur between the two. Water will move from the protein solution to the reservoir causing the precipitant and protein concentration to increase, in some cases to a supersaturated state. The screening condition will usually lead to either precipitation of protein or a clear drop, but very occasionally the supersaturated protein solution will nucleate and crystallise – *some luck is needed*.

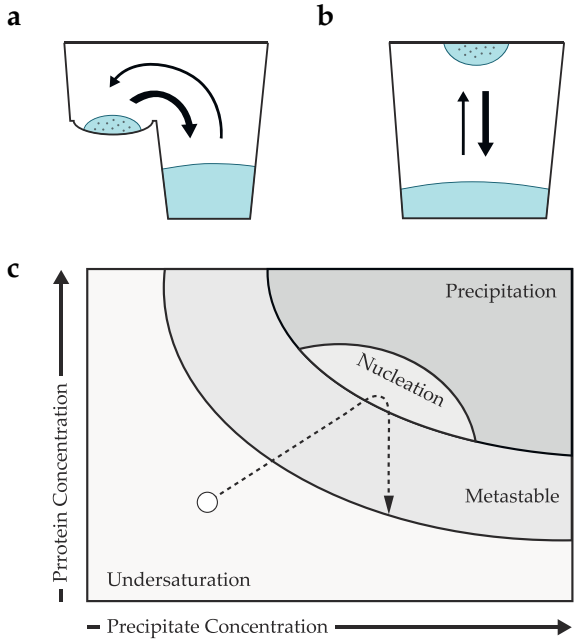


Figure 2.1: Vapour Diffusion
Proteins can be crystallised by vapour diffusion in either a **(a)** sitting-drop or **(b)** hanging-drop format. Water vapour will diffuse between the drop and the reservoir with a net flow toward the reservoir causing the protein and precipitant concentration to rise. **(c)** A schematic of the crystallisation phase diagram. The dashed-line depicts the path taken during a successful crystallisation experiment.

The simplest unit within a crystal is called the asymmetric unit which can combine to form a repeating unit cell. The asymmetric unit is duplicated, translated and rotated as described by specific symmetry operations to generate the unit cell. Mirror symmetry is also possible in small molecule crystals but does not occur in protein crystal due to the presence of chiral centres of amino acids. The unit cell can produce a crystal when repeated and translated, which in turn leads to a repeating arrangement of atoms and planes. The unit cell is a parallelepiped that can be defined by three vectors, a, b, and c, whose angles between each vector are α, β, and γ (Figure 2.2).

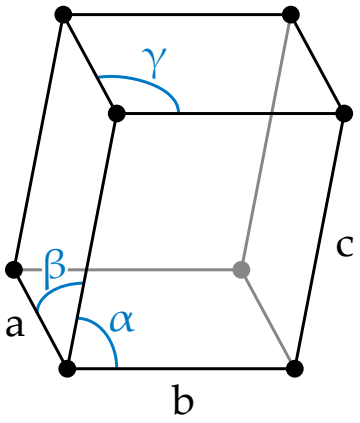


Figure 2.2: Unit Cell
Crystals are made up of repeating units termed unit cells which are parallelepipeds defined by their lengths, a, b, c and the angles between, α, β, γ.

There are seven types of crystal systems with each specifying allowed vector lengths and angles of the unit cell (Table 2.1). The vector lengths and angles are a result of the underlying symmetry and can be used to indicate the point group of a crystal. The point group describes the rotational symmetry of the asymmetric unit along the X, Y, and Z axis. Further to the point group there exist screw axis, denoted by a subscript number (*e.g.* $P2_1$) which describes a combined rotation and translation. Combining all possible symmetries (point groups and screw axes) there exist 230 difference space groups. However as proteins are chiral molecules and mirror symmetries are not available, protein crystals are limited to one of a possible 65 different space groups.

Table 2.1: Crystal Systems

The names of different crystal systems and their possible point groups given the possible values of the unit cell lengths a, b, c, and angles α, β, γ. Point groups are given in Hermann-Mauguin notation (Hermann, 1928; Mauguin, 1931).

System	Angles (°)	Cell-lengths	Point group
Triclinic	$\alpha \neq \beta \neq \gamma \neq 90$	$a \neq b \neq c$	1
Monoclinic	$\alpha = \gamma = 90$ and $\beta \neq 90$	$a \neq b \neq c$	2
Orthorhombic	$\alpha = \beta = \gamma = 90$	$a \neq b \neq c$	222
Tetragonal	$\alpha = \beta = \gamma = 90$	$a = b \neq c$	4, 422
Trigonal	$\alpha = \beta = 90$ and $\gamma = 120$	$a = b \neq c$	3, 32
Hexagonal	$\alpha = \beta = 90$ and $\gamma = 120$	$a = b \neq c$	6, 622
Cubic	$\alpha = \beta = \gamma = 90$	$a = b = c$	23, 432

2.3 X-ray Diffraction

To determine a protein crystal structure is to determine the arrangement of atoms within the crystal. The periodic and repeating nature of a crystal can be exploited with the use of X-rays to uncover the underlying atomic arrangement. When subjected to a intense beam of X-ray photons, the repeating planes within the crystal, that are themselves a result of the underlying repeating atomic arrangement, cause the photons to diffract and exit the crystal at a different angle to the incoming angle. During diffraction the incoming photons will simultaneously interact with all atoms of a crystallographic plane and its phase will be affected upon diffraction. If the phase angle of the reflected wave constructively interferes with the phase angle of waves reflected by equivalent crystallographic planes then the intensity of the diffracted wave can be measured. Bragg's law (Equation 2.1) describes the condition on the scattering angle (θ) at which constructive interference of the diffracted waves is strongest depending on the incoming wavelength, λ, and the interplanar spacing, d (Bragg and Bragg, 1913). Under these specific conditions of constructive interference, the properties of the diffracted X-ray photons can be used to calculate the internal crystallographic arrangement.

$$n\lambda = 2d \sin\theta \tag{2.1}$$

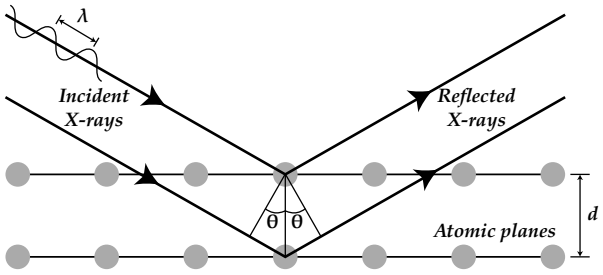


Figure 2.3: Bragg's Law When a crystal diffracts X-rays, peaks are observed when the angle of incident X-rays equals the angle of reflected X-rays and when the path length difference, $2d \sin\theta$, is equal to n wavelengths.

2.4 X-ray Diffraction Data Collection

The interaction of X-rays with matter results in many phenomena including absorption, fluorescence, and scattering. The aim of a diffraction experiment is to accurately determine the intensities of the Bragg reflections from a crystal which result from X-ray scattering. When collecting diffraction data, there are a number of parameters that can be adjusted. In the past it was typical to collect X-ray diffraction data 'in-house' on an X-ray source such as a rotating copper anode. In this case, the X-ray source consists of an electron beam from a cathode hitting the anode which causes atoms within the anode to lose an inner-shell electron. A higher energy outer-shell electron is then able to fill its place, emitting an X-ray photon in the process. However, it is now routine to screen crystals and collect X-ray diffraction data at a synchrotron source due to their increased availability. Synchrotrons have the enormous benefit of producing highly brilliant X-ray light, enabling fast and high-resolution data collection. Synchrotrons produce X-rays by first accelerating electrons to high energies around a 'storage-ring'. The 'storage-ring' is not actually circular and is instead a polygon consisting of many straight sides connected by bending magnets. As the electron beam changes direction through each bending magnet on the storage ring, an intense beam of radiation is released. Straight sections of the synchrotron storage-ring can also be used to produce beams of radiation through the use of an undulator. An undulator consists of a linear series of magnets each with opposing poles that cause the electron beam to follow the magnetic field in a wave, releasing radiation in the process. The beams of X-ray radiation produced at a synchrotron is then focussed and the wavelength tuned for experimental use on beamlines.

On the beamline, a crystal is mounted on a goniometer and centred so that the X-ray beam passes directly through. A goniometer is a device that allows precise positioning

of the crystal – modern goniometers allow up to six axes of movement. To collect X-ray diffraction data, the crystal is typically centred with respect to the beam and rotated around one axis, while images of diffracted photos recorded at regular intervals (Figure 2.4). On modern, high-brilliance beamlines with fast pixel-array detectors it is now common to collect fine-sliced and low-intensity data. Fine-slicing refers to an X-ray data collection strategy where small oscillations (typically around 0.1°) of the goniometer are used to collect a greater number of diffraction frames, leading to an increase in data quality (Mueller *et al.*, 2012). Instead of full-reflections being recorded on a single frame, fine-slicing results in the recording of partial reflections which can give a much greater background separation and thus a better measurement of the true reflection intensity. Pixel-array detectors also have a much greater signal-to-noise ratio and can detect single diffracted photons (Broennimann *et al.*, 2006). Inclusion of weak-intensity X-ray diffraction data has been shown to not only improve the quality of crystal structures (Wang and Wing, 2014) but also allow collection of near radiation-damage free data (Carugo and Djinović, 2005).

Many macromolecular X-ray crystallography beamlines at synchrotrons are setup to have a fixed energy near to the selenium k-edge (12.6578 keV). While ideal for solving selenomethionine labelled proteins by Se-SAD or MAD, this energy can also exploit the anomalous scattering of zinc, which is natively present among many proteins (Figure 2.5).

2.5 Cryo-cooling Crystals

Before x-ray data collection, it is now common practice to cryo-cool the crystals to a temperature of 100 K to mitigate some effects caused by X-ray induced radiation damage. The cryo-cooling of crystals comes with its own set of limitations. The collection of diffraction data at room temperature is increasingly avoided, however in some rare cases cryo-cooling can disrupt the crystal lattice and room temperature collection is necessary. Another problem with the cooler temperatures is ice-formation. Typically crystals are mounted onto a loop whereby the crystal is surrounded by a small amount of the crystallisation solution before plunging into liquid nitrogen. This crystallisation solution can form ice crystals which can both cause disruption to the crystal lattice and/or the ice crystals themselves can diffract X-ray photons, producing

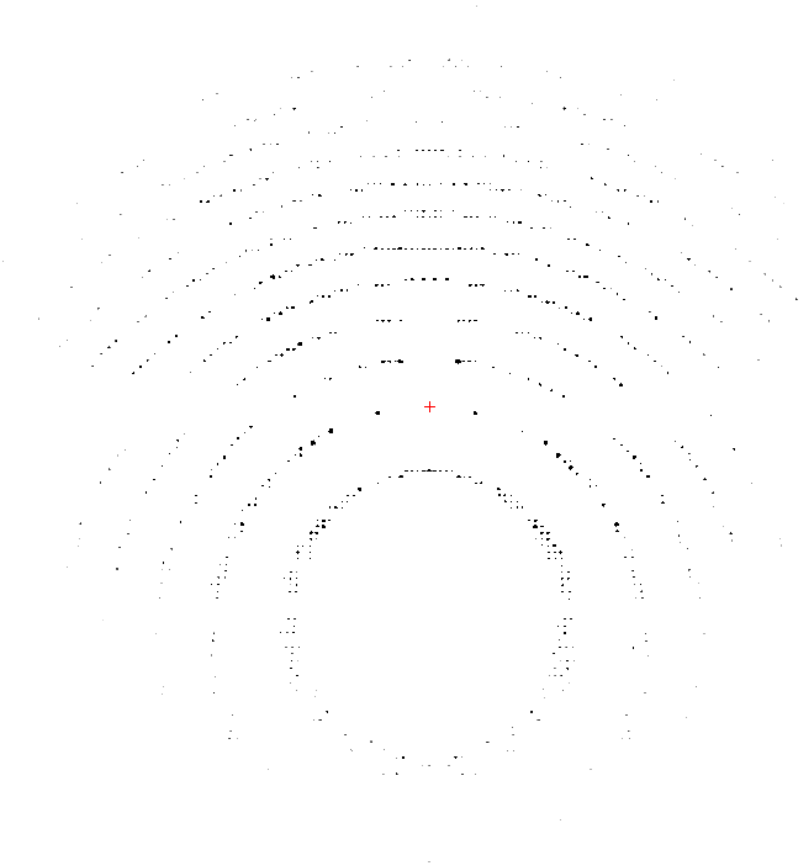


Figure 2.4: Diffraction An example diffraction image from a crystal of botulinum neurotoxin H_C/A1. Each spot (or reflection) is a result of diffraction that satisfies Bragg's law. The red cross indicates the location of the beam centre, with crystals diffracting to higher resolutions, the further their spots from the centre.

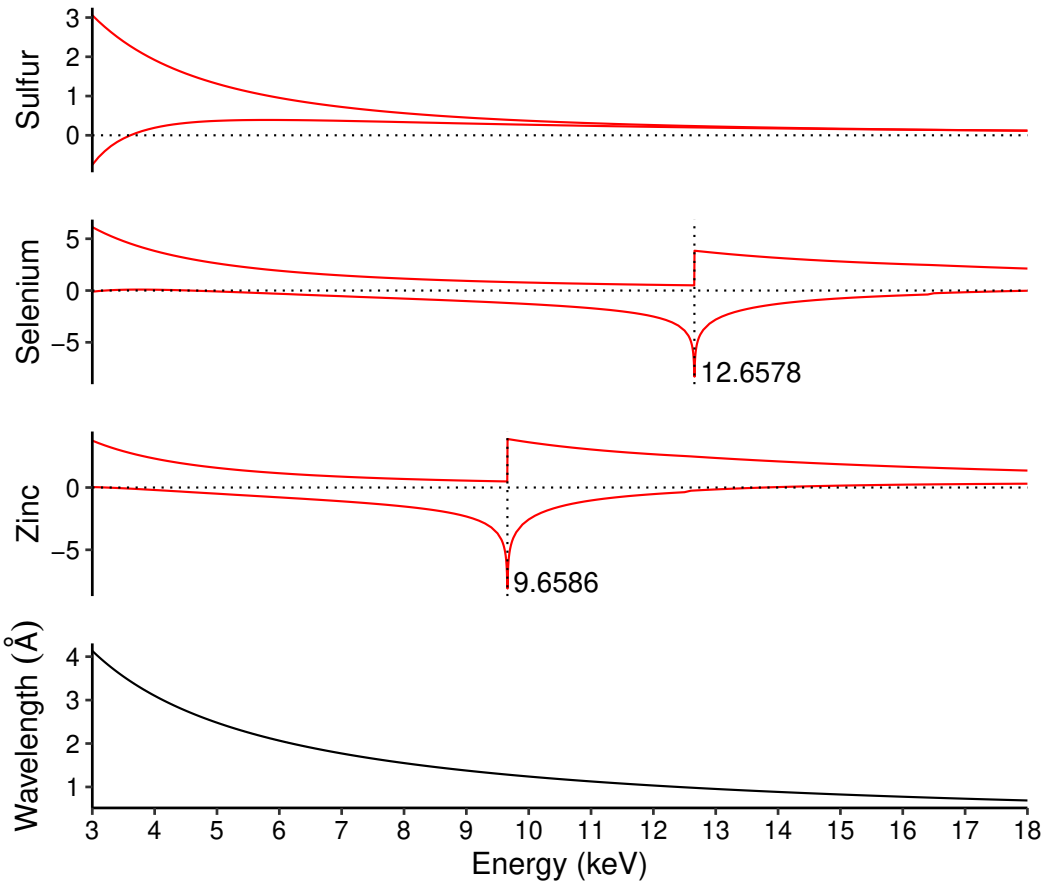


Figure 2.5: X-ray Absorption and Scattering Factors. Theoretical absorption scattering factors for sulfur, selenium and zinc over the energy range available accross MX beamlines at Diamond Light Source. Energy can be readily converted to wavelength using $E = \frac{hc}{\lambda}$.

unwanted artefacts in recorded data (Parkhurst *et al.*, 2017). Methods to avoid ice formation aim to increase the cooling rate through fast quenching into liquid nitrogen and cryo-protecting the crystal through the addition of chemicals which disrupt ice formation such as glycerol (Warkentin and Thorne, 2007). At cooler temperatures, the radiation damage caused to the crystal by the X-ray beam is greatly reduced (Low *et al.*, 1966). Radiation damage is a major concern, especially when collecting data using a high brilliance x-ray source such as a synchrotron. X-rays can cause damage to protein crystals by producing reactive species, for example H^* from solvent, directly damaging covalent bonds or through thermal energy dissipation (Carugo and Djinović, 2005). By reducing the temperature of the crystal and surrounding mother liquor to 100 K, the spread of reactive species within the crystal is reduced and thermal damage becomes less of a problem.

2.6 X-ray Data Reduction

2.6.1 Indexing and Integration

Following data collection the location of each reflection on the diffraction image is recorded. This is then converted to a Miller Index, which related the reflection to a specific plane that produced the diffracted photons, and the Laue group of the diffracting crystal can then be determined following the conditions set out in Table 2.1. The data are then integrated across images to calculate the intensity of each measured reflection. The intensity of a given reflection is the result of subtracting the background intensity from the recorded peak intensity. Following integration, the full space group and be inferred from systematic absences within the list of reflections. The space group can be determined with a high confidence in most cases but it can not be confirmed until the structure has been solved. Improvements in software and automation have led to this once time consuming process now only taking a matter of minutes following collection of a full X-ray dataset.

2.6.2 Scaling and Merging

After integration, the data are scaled to correct for some experimental variables during data collection. At modern synchrotrons, the X-ray beams are extremely stable but still can have slight variations of intensity during the course of an X-ray diffraction experiment. The process of scaling compares integrated reflections between images and adjusts their intensities to account for some of the non-uniformity. Following scaling the data are then merged to determine the average intensity for each of the Miller indices and their associated standard deviation. During merging, a number of statistics are used to assess the quality of the collected data across bins or ranges of resolution. By recording multiple observations of the same reflection, one can determine to what extent they agree with one another and if there are any outliers. Most scaling programs will remove weak or poor data which is classed as an outlier with respect to other observations. An early method to determine the quality of collected data was by using the R_{merge} statistic.

$$R_{merge} = \frac{\sum_{hkl} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{n} I_i(hkl)} \tag{2.2}$$

The R_{merge} sums for all reflections: the difference between the intensity of a measured reflection and the average of all intensities for the same reflection. Lower R_{merge} values typically indicate better agreement between all observations of the same reflections and thus better data overall. When using data that contains a high multiplicity – that is, many observations of the same reflections – such as those merged from multiple crystals or datasets, the R_{merge} quantity is unable to accurately reflect the data quality. R_{merge} will increase with multiplicity even when the included data is of high quality. A new quality indicator, R_{meas} (Equation 2.3), was introduced by Diederichs and Karplus (1997) which is a modification of R_{merge} that accounts for multiplicity. R_{meas} is obtained by adjusting R_{merge} by a value of $\sqrt{\frac{n}{n-1}}$ where n is the multiplicity.

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{n} I_i(hkl)} \tag{2.3}$$

R_{meas} is always greater than R_{merge} but as the data multiplicity increases toward infinity, they will become equal. To realise the importance of a quality indicator for datasets of high multiplicity, R_{pim} was created. R_{pim} is used to give an indication of the data precision at high multiplicity.

$$R_{pim} = \frac{\sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{n} I_i(hkl)} \tag{2.4}$$

Another method to quantify the quality of unmerged data is by determining the correlation coefficient between two half-datasets ($CC_{1/2}$). To calculate the $CC_{1/2}$, reflections are first be binned by resolution and then split into two half-datasets. A Pearson's correlation coefficient is then used to calculate the agreement (Equation 2.5). At higher resolutions the spot intensity decreases and the data becomes increasingly noisy thus, the correlation between half-datasets decreases. A $CC_{1/2}$ calculated from un-binned data is not helpful as low-resolution reflections tend to overpower the calculation.

$$CC_{1/2} = \frac{\sum (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum (a_i - \bar{a})^2} \sqrt{\sum (b_i - \bar{b})^2}} \tag{2.5}$$

Assigning reflections randomly to one of the half-datasets will occasionally result in a situation where you are left with a lone reflection. For example, when a unique reflection is observed an odd number of times. In such a case, a random number can be assigned as its pair in the opposing half-dataset (Prof. Kay Diederichs, personal communication, 25th May 2017). Including random data in this way can result in different outcomes if the $CC_{1/2}$ is calculated more than once. To overcome this problem some programs implement a loop to calculate the value multiple times and then report the average value. An alternative approach suggested by Assmann *et al.* (2016) is to avoid splitting reflections into half-datasets entirely, named the 'σ-τ method' (Equation 2.6).

$$CC_{1/2} = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma_\epsilon^2} \tag{2.6}$$

Each of these metrics is used by a crystallographer to assess the recorded data quality during reduction of the diffraction images into merged data. Abnormal values, such as a particularly high R_{meas} for low multiplicity data indicates problems which may cause issues during structure solution. Many data problems are the result of poor data collection strategies or data artefacts from ice-rings but some are simply the result of the intrinsic properties of a particular crystal. These metrics are also used to determine where to decide a resolution cut-off or in other words, deciding when the data simply becomes noise. Some publications have suggested that data in the highest resolution shell should be included as long as the $CC_{1/2}$ is greater than 0.1 (Karplus and Diederichs, 2015) and indeed there seems to be little negative impact to the final model from including this weak data (Evans and Murshudov, 2013).

2.7 Solving a Crystal Structure

As mentioned in the previous sections, recorded X-ray diffraction data contains the location of each reflection on the detector and thus its Miller index, and the intensity of each reflection. A third parameter is needed to calculate the underlying atomic structure and that is the phase angle of each reflection. There is no direct way to record the incoming phase and so this information is lost. To convert the recorded data into electron density, into which a molecular model can be built, the information is Fourier transformed using the electron density equation (Equation 2.7).

$$\rho_{(xyz)} = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{(hkl)}| \exp[i\phi(hkl)] \exp[-2\pi i(hx + ky + lz)] \tag{2.7}$$

Here, the structure factor amplitute, $F_{(hkl)}$, is directly proportional to to the square root of the intensity of each reflection, but the phase of each reflection, $\phi(hkl)$, must be known or calculated. In an X-ray diffraction experiment there are three ways to calculate or estimate the phase of each reflection – direct methods, experimental phasing or molecular replacement – with the latter of these methods most commonly used in modern protein crystallography due to the complexity of protein structures. Direct methods relies on prior knowledge that correct phases will result in a positive electron density map which is evenly distributed within the unit cell. It involves

multiple trials whereby the phases are estimated and in a small number of cases, where enough of the phases are correct, the structure can be solved (Hauptman, 1997).

Molecular replacement exploits the similarity between proteins of similar sequence and thus presumably similar structure to estimate the unmeasured X-ray diffraction phases (Rossmann, 1990). First the Patterson function is calculated from both the experimental data and the existing molecular model. A Patterson function is derived from the electron density equation but with the phase contribution set to zero, resulting in Equation 2.8 (Patterson, 1935).

$$\rho_{(xyz)} = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{(hkl)}|^2 \exp[-2\pi i(hx + ky + lz)] \tag{2.8}$$

This results in two 3-dimensional representations (maps) of the intermolecular vectors and intramolecular vectors within the unit cell, which appear as peaks. The vectors can be measured and atomic positions calculated. The search model is then rotated and a new map calculated until there are peaks which match the experimental data. Following a rotation search, a translation search is carried out in the same manner. Once a highly likely solution has been found, the calculated phases from the search model can be included with the $F_{(hkl)}$ of the experimental data and electron density can be determined (Rossmann, 1990). Molecular replacement is the most common approach to solving crystal structures now as only partial knowledge of the protein fold is needed (McCoy, 2007). Modern software is further pushing the boundaries of molecular replacement and in the case of high resolution data, a small amount of protein fragments such as α-helices can suffice (Sammito *et al.*, 2014).

2.8 Refinement of the Atomic Model

The starting model must be manually built or edited into the density, in real space, following phasing, be it experimentally or from molecular replacement. The two most common types of electron density map are $2F_o - F_c$ and $F_o - F_c$ where F_o is the observed structure factor amplitudes and F_c are the amplitudes calculated from the atomic model. The $2F_o - F_c$ map is what one would usually build the atomic model into, it is the best available estimate for the true electron density. The $F_o - F_c$ map is

known as a Fourier difference map, simply because it displays differences between the model and the crystal – positive differences indicate the model does not account for some electron density and negative differences are the opposite. Generally, the electron density toward the start of model building is poor as the phases need to be improved from the model. Some parts of the protein structure may be poorly ordered within the crystal, such as solvent-exposed side-chains and loops, and in these cases it can be impossible to accurately generate parts of the model. After a model has been built or modified to fit the density it can be refined to minimise the differences between the F_o and F_c. In turn the improved atomic coordinates can also improve the structure factor phases used to calculate the electron density. Modern refinement software uses the maximum likelihood method to refine the protein model (Murshudov *et al.*, 1997; Afonine *et al.*, 2012). The refinement software is provided with parameters including the structure factor amplitudes, phases from calculated or experimental methods, and model coordinates. In addition, most software makes use of prior knowledge such as empirical geometry restraints (Engh and Huber, 1991).

The resolution is also an importance factor when modelling and refining a protein structure. There exists a 'sweet-spot' with regard to the resolution of the determined protein structure. At low resolution (less than 3 Å), one cannot accurately position side-chains, α-helices can resemble single tubes of density, and model building is very challenging. At medium resolutions (between 3 Å and 1.6 Å), most side-chains are apparent, there are a reasonable number of water molecules, and model-building is generally very easy. At very high resolutions (better than 1.6 Å), it is common to observe more than one conformation of side-chain, backbone, ligand, etc., and many waters can be found, all of which can become extremely difficult to understand and model. Of course given the choice, no real crystallographer will opt for anything other than the very best resolution possible, but even the prettiest of crystals give no guarantee of high resolution.

2.8.1 B-factor Refinement

Every atom described in the model contains an additional factor which is used to describe the displacement of the atom from mean position of the atom – this is usually referred to as the temperature factor or B-factor. No atoms within a crystal are static, even at very low temperatures, thus the B-factor is an essential term in refinement since the goal is to improve agreement between the model and the measured reflections. Most protein models contain isotropic B-factors, which describe the displacement of an individual atom using a single term. At higher resolutions, and with many more reflections, it is possible to better describe the atomic displacement using anisotropic B-factors which normally utilise six parameters instead of the usual one. The anisotropic displacement parameters (ADPs) are capable of describing the atomic displacement in three dimensions which, at higher resolutions, can better describe the crystallographic data (Figure 2.6).

Refinement is an iterative process between manual editing of the model and automated refinement. The automated refinement and minimisation of the distance between the model and data cannot fix major errors, nor can it fix minor errors such as incorrectly orientated side-chains. There is still no replacement for the human crystallographer who can refine to the best of their ability and experience, but even so — and as first quoted by Philip R Evans — most crystallographers refine *ad tedium*.

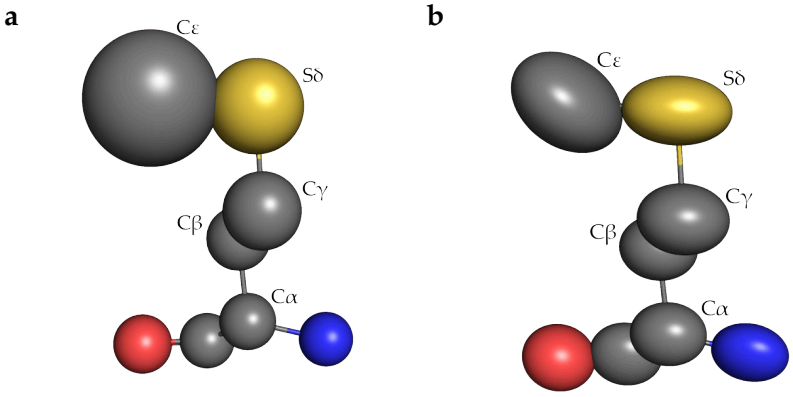


Figure 2.6: Isotropic and Anisotropic B-Factors
Graphic representation of isotropic **(a)** and anisotropic **(b)** B-factors of atoms from a methionine amino acid.

2.9 Validation of the Atomic Model

2.9.1 Model bias and R_{free}

When refining a model against crystallographic data, it is useful to quantify their agreement. The R_{work} (also referred to as the R value) is a quantity which is produced by calculating the sum of the difference between the observed structure factors and the calculated structure factors, and dividing by the sum of the observed structure factors (Equation 2.9). As the model is built and refined it will have a better agreement to the recorded data and therefore the R_{work} will approach zero. In practice, the R_{work} will never reach zero owing to noise in the data and the limited resolution available. During refinement the R_{work} will generally decrease until the model cannot be improved further. It is possible to artificially reduce the R_{work} below its true value through over-fitting the model to the data, where atoms are modelled into noise.

$$R = \frac{\sum_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)|}{\sum_{hkl} |F_{obs}(hkl)|} \tag{2.9}$$

To reduce the chance of over-fitting, some reflections can be kept aside – *free reflections* – and not used for the refinement of the model, instead they can remain independent and used to calculate an unbiased statistic (Brünger, 1992). Equation 2.9 can be reused with the free reflections to calculate the R_{free} value. If a model has been refined with the *free* reflections excluded from the total data then model should be unbiased with respect to the *free* reflections. Generally, a small percentage of reflections are chosen at random from a list of unique reflections while ensuring an equal distribution across each resolution bin. The error of R_{free} is approximately equal to $R_{free}/(2n)^{1/2}$ (Tickle *et al.*, 2000). Thus, for an R_{free} value of 24%[†], selecting 2000 *free* reflections would give an approximate error of 0.38%. During model building and refinement, both the R_{work} and R_{free} should decrease in response to an improved fit to the experimental data. An increase in R_{free} is strong indication of over-fitting and should be avoided.

[†]The mean R_{free} across all structures in the PDB is 24% ($n = 116570$, correct as of 04 Jan 2018)

2.9.2 Ramachandran Plot

Another method for quickly ensuring that protein model geometry is reasonable is to produce a Ramachandran plot (Ramachandran *et al.*, 1963). The Ramachandran plot can be used to visualise the allowed angles for backbone dihedral angles (ϕ and ψ, Figure 2.7,2.8) within a protein structure. By analysing existing high resolution structures, one can produce confident approximations of allowed angles for ϕ and ψ and use these to validate a new protein model. Modern Ramachandran plots use data from thousands of structures which have been filtered to ensure only models regarded as good are used as a reference (Read *et al.*, 2011). Of course, every method of validation has its limitations and many proteins contain a small percentage of residues with backbone dihedral angles that sit outside of the allowed regions.

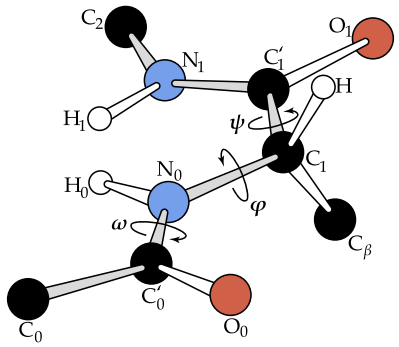


Figure 2.7: Ramachandran angles.
Representation of two peptide residues indicating locations of ϕ, ψ and ω angles. Image redrawn and modified from Ramachandran *et al.*, 1963.

In addition to bond angles, the bond lengths can also be used as a good indicator of model quality. After analysis of many small molecule crystal structures as a reference, Engh and Huber (1991) released the first set bond length parameters for use in protein structure refinement. As the reference parameters are derived from atomic or near-atomic resolution small molecule crystal structures, all protein model (which are generally produced from lower resolution data) should fit without many deviations. There is some still some debate about how closely to follow ideal bond-angle and bond-length values and in practice, following the electron density should always be the primary method for building your model. Large deviation from ideal however should cause concern and would need strong justification (Wlodawer *et al.*, 2013).

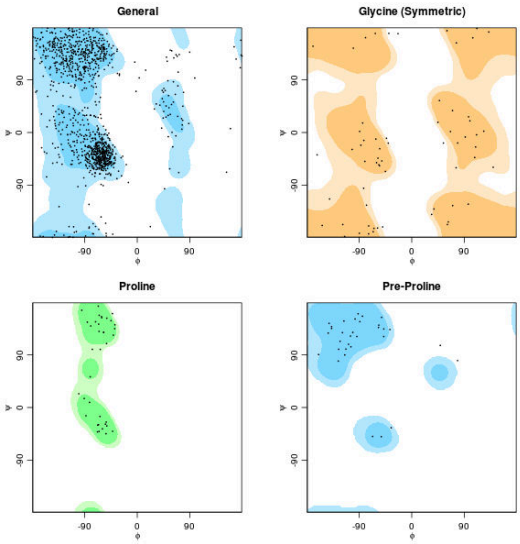


Figure 2.8: Ramachandran Plot
An example Ramachandran plot showing the residues from the full-length crystal structure of BoNT/A1 (Lacy *et al.*, 1998).

2.10 Summary

X-ray crystallography is an incredibly powerful technique for studying protein structure. The background and theory presented here should serve as a general overview and aid in understanding both the reasoning and approach followed for some of the presented data in the following chapters.

CHAPTER 3

Structural Studies of BoNT/A H_C Domains

3.1 Introduction

Botulinum neurotoxins (BoNTs) comprise a family of proteins produced predominantly by *Clostridium botulinum* (Schiavo *et al.*, 2000; Rossetto *et al.*, 2014). There are seven distinct serotypes termed BoNT/A through BoNT/G that cause the deadly condition botulism in vertebrates. Of these serotypes, /A, /B, /E and /F are most commonly associated with human botulism (Coffield *et al.*, 1997). Each BoNT serotype, for example BoNT/A, can be further categorised into subtypes based on amino acid sequence identity. For BoNT/A there are currently eight known subtypes (BoNT/A1–/A8) which share between 84% and 97% sequence identity (Table 3.1). The most widely used commercially available therapeutic BoNT is the subtype BoNT/A1 (Fonfria *et al.*, 2018). One other serotype, BoNT/B is also commercially available although it is less potent than BoNT/A formulations due to a residue difference within the protein receptor for BoNT/B, human synaptotagmin II, compared with its murine counterpart (Peng *et al.*, 2012; Strotmeier *et al.*, 2012; Tao *et al.*, 2017). Other BoNT serotypes have been explored for use as therapeutic molecules but much less work has focussed on differences between subtypes from within BoNT serotypes.

Table 3.1: H_C/A subtype identities.
Percentage identities are given for full-length and H_C domain sequence alignments of each BoNT/A subtype[†]. For the H_C alignments, sequences were aligned to BoNT/A1 residues K870 to L1296.

	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	
A1	–	87.29	86.82	91.55	93.90	90.61	91.78	87.79	
A2	90.04	–	98.83	88.47	89.65	90.12	90.35	93.43	
A3	84.66	93.19	–	88.24	88.94	89.65	89.88	92.72	
A4	89.19	88.12	84.29	–	86.85	85.92	85.92	90.14	
A5	97.15	90.50	85.21	87.34	–	93.43	92.72	89.91	
A6	95.68	91.74	86.29	87.72	95.91	–	91.08	87.32	
A7	93.75	89.81	84.90	86.64	94.37	92.98	–	89.67	
A8	93.36	93.44	87.69	88.81	93.60	93.06	91.36	–	
	Full-length identity								**H_C identity**

[†] Uniprot accession codes for BoNT/A1 to /A8 are A5HZZ9 (Betley *et al.*, 1989; Gimenez and DasGupta, 1993; Sebaihia *et al.*, 2007; Smith *et al.*, 2007), Q45894 (Dover *et al.*, 2013), Q3LRX9 (Jacobson *et al.*, 2008), Q3LRX8 (Jacobson *et al.*, 2008), C7BEA8 (Dover *et al.*, 2009), C9WWY7 (Luquez *et al.*, 2009), K4LN57 (Mazuet *et al.*, 2012) and A0A0A7PDB7 (Kull *et al.*, 2015), respectively.

Despite their high sequence identity and similarity, it is important to determine the

differences between BoNT subtypes and their impact on intoxication as these may be useful therapeutically. The sequences of H_C domains from BoNT/A subtypes align well and all sequences share at least 80% identity with one another (Figure 3.1). Alignment of the known sequences for BoNT/A subtypes identifies many amino acid variations overlapping or in close proximity to residues involved in receptor binding (Figure 3.1). The regions of the H_C domain involved in receptor binding have been characterised to various extents. While the SxWY motif within the ganglioside binding site (GBS) is conserved, many of the residues around this site show high variability. Some BoNT/A subtypes have been reported to exhibit different properties to BoNT/A1 such as onset of symptoms and duration of action. For example, BoNT/A3 has been shown to be less effectively neutralised by anti-BoNT/A1 antibodies whilst also inducing symptoms of intoxication in mice that are distinct from BoNT/A1 (Mazuet *et al.*, 2010; Tepp *et al.*, 2012). It has been suggested that these differences are due to structural differences within the H_C of BoNT/A3 (H_C/A3).

Between subtypes from the same serotype there can be distinct intoxication characteristics. A long-term study of BoNT activity in cultured neuronal cells and in a mouse model found that while BoNT/A1, /2, /4, and /5 have a similar duration of action (over 10 months), the duration of BoNT/A3 is significantly shorter (3 months) (Pellett *et al.*, 2015). The reduced duration of action may be due to differences in the LC alone. A hybrid toxin containing the LC/A1 and LH_N/A3 maintains a long duration of action whereas a hybrid containing LC/A3 and LH_N/A1 results in a much shorter duration (Pellett *et al.*, 2018). Interestingly, it has also been determined that BoNT/A2 displayed a much faster onset of paralysis compared to other BoNT/A subtypes (Pellett *et al.*, 2015). The rate limiting step, determining the onset of durations is most probably a result of the H_C and H_N region of the BoNT which is responsible for binding and translocation of the LC (Pellett *et al.*, 2018). Other evidence of differences between subtypes intoxication was identified by Pellett *et al.* (2015), who discovered that BoNT/A4 is ~1000-fold less potent than BoNT/A1, the reason for this is as yet unknown. Such examples highlight the need to investigate BoNT subtypes and their function. Due to the modular nature of the BoNT structure, it may prove useful in future to produce hybrid BoNTs containing domains from different subtypes to tailor the intoxication properties.

In addition to the BoNT/A subtypes, a recently discovered BoNT hybrid toxin (BoN-

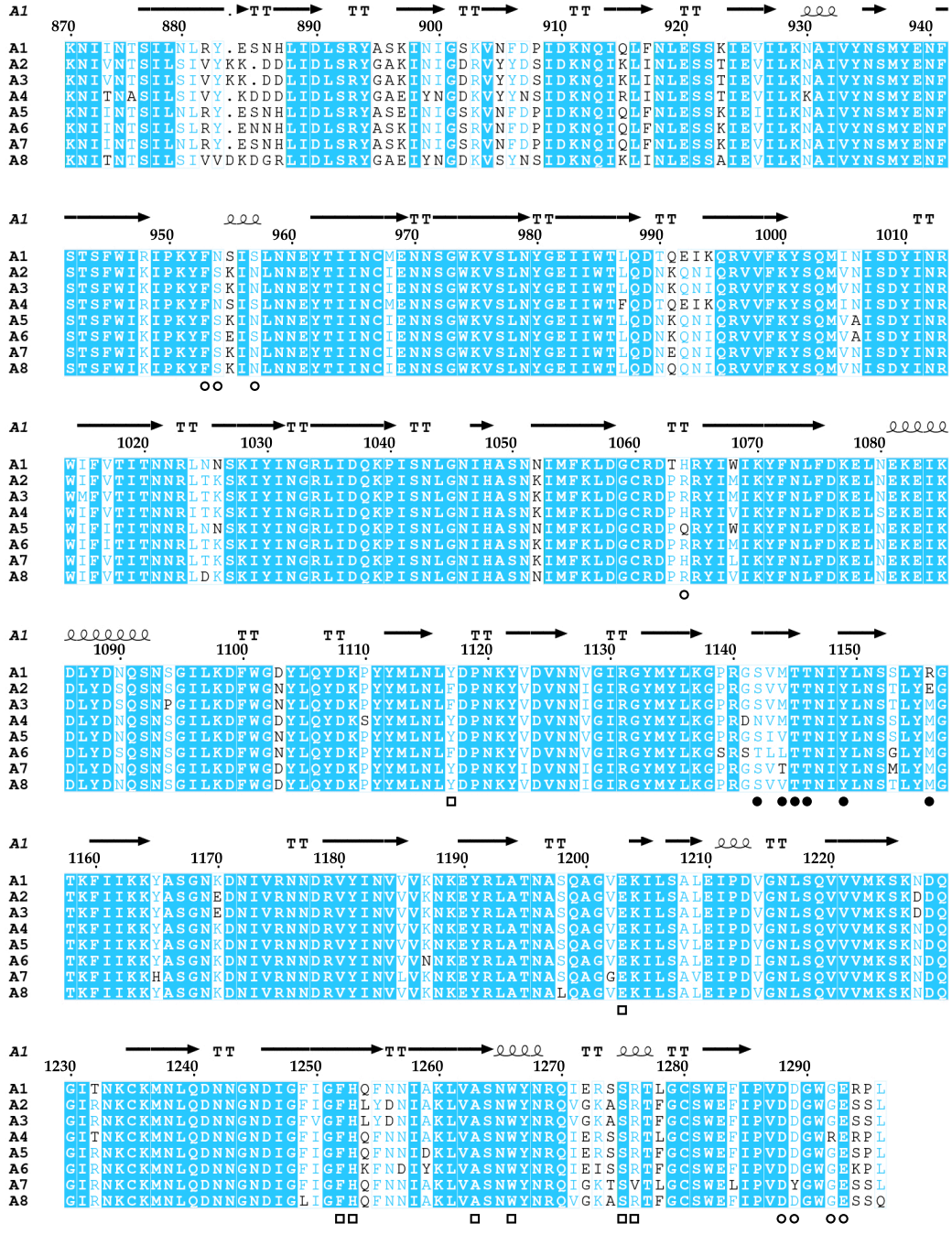


Figure 3.1: H_C/A Primary Sequence Alignment

Residues that interact with receptors are identified with a symbol below the sequence: ganglioside □; SV2–backbone ●; SV2–glycan ○. Primary sequences of H_C/A subtypes aligned using Clustal Omega and figure produced using ESPript (Robert and Gouet, 2014).

T/FA) has some interesting features (Barash and Arnon, 2014; Dover *et al.*, 2014). It was found to contain a H_C domain similar to that of BoNT/A1 and a LC and H_N domain similar to BoNT/F5 (Gonzalez-Escalona *et al.*, 2014). As the H_C is similar to BoNT/A1, it likely shares the same receptors and could provide valuable information when exploring the engineering of novel therapeutics.

3.2 Methods

All materials were purchased from Sigma Aldrich unless otherwise specified.

3.2.1 General - SDS-PAGE

SDS-PAGE was carried out using precast 4–12% gradient gels. 5 µL of 5x protein sample buffer (10% w/v SDS, 50% v/v Glycerol, 250 mM Tris-HCl and 0.5% w/v bromophenol blue dye, pH 6.8) was added to 15 µL of the protein sample to be analysed. Each sample was heated to 95°C for 2 minutes prior to loading. SDS-PAGE gels were run at 200 V for 45 minutes in a solution of TruPAGE TEA-Tricine SDS buffer. The gels were removed from electrophoresis, washed in water and stained using InstantBlue protein stain (Expedeon) for 1 hour. Gels were destained in water for 1 hour.

3.2.2 General - Microbiological Techniques

All experiments were performed using standard aseptic technique under flame. Media and glassware used for bacterial culture was autoclaved prior to use at 121°C, 100 kPa for 15 min. All buffers were filtered through a 0.2 µm membrane. The pH of each buffer was adjusted using 5 M NaOH or concentrated HCl where appropriate and measured at room temperature using a pH meter calibrated with standard buffers at pH 7.0, pH 4.0, and pH 10.0.

3.2.3 Cloning of H_C/A and H_C/FA Expression Constructs

BoNT/A1 to /A6, and /FA H_C domains were sub-cloned from constructs containing the respective full-length *bont* sequences (provided by Ipsen Bioinnovation Ltd.). The region corresponding to the H_C was amplified by PCR using the primers listed in Table 3.2. The forward primers encode an additional N-terminal 6xHistidine-tag for protein purification. Amplified DNA fragments were flanked by an N-terminal NdeI and a C-terminal HindIII restriction site. The DNA was cleaved by restriction digestion

using both NdeI and HindIII (NEB) at 37°C for two hours. Digested DNA was purified using a PCR clean up kit (Wizard, Promega). The vector, pJ401 (DNA2.0), was also digested using NdeI and HindIII and purified by gel purification. Inserts were ligated into digested pJ401 using T4 DNA ligase (NEB) at 22°C for 16 hours. Ligated vectors were transformed into *E. coli* TOP10 cells (Invitrogen) using the following method. A 50 µL aliquot of frozen cells were thawed on ice for 10 min. 2 µL DNA from the ligation was added and the cells returned to ice for 15 min. Cells were heat-shocked at 42°C for 30 seconds before returning to ice for a further 5 minutes. 150 µL SOC media (2% w/v tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose) was added to each aliquot and cells were grown at 37°C for 1 hour while shaking at 225 rpm. 100 µL of cell culture was used to inoculate a pre-warmed LB (Miller) agar plate containing 50 µg mL^{-1} kanamycin sulfate. Plates were incubated at 37°C overnight. Individual colonies from each ligation plate were used to inoculate 10 mL LB (Miller) media and the culture was grown for 16 hours at 37°C while shaking at 225 rpm. Plasmid DNA was extracted from each culture using a miniprep kit (Wizard, Promega) and sent for DNA sequencing (Eurofins Genomics, Germany) to confirm successful cloning.

Table 3.2: H_C/A Cloning Primers

Name	Sequence
H_CA1_F	AACATATGCATCATCACCATCACCACAAAAACATCATCAATACTAGCATTCTGAACCTG
H_CA1_R	AAAAGCTTTTACAGCGGACGTTCGCCCCA
H_CA2_F	GCAGCATATGCATCATCACCATCACCACAAGAACATCGTCAATACGTCTATCC
H_CA2_R	GCCTCAAGCTTTTACAGGCTAGACTCACCCCA
H_CA3_F	GCAGCATATGCATCATCACCATCACCACAAAAACATCGTCAATACGTCTATTCTG
H_CA3_R	GCCTCAAGCTTTTACAGCGAGCTTTCACCC
H_CA4_F	GCAGCATATGCATCATCACCATCACCACAAAAACATTACCAATGCGTCCATC
H_CA4_R	GCCTCAAGCTTTTACAACGGACGTTCGCG
H_CA5_F	GCAGCATATGCATCATCACCATCACCACAAAAACATCATTAACACCAGCATCT
H_CA5_R	GCCTCAAGCTTTTACAGCGGGCTCTCG
H_CA6_F	GCAGCATATGCATCATCACCATCACCACAAGAACATTATCAACACGAGCATC
H_CA6_R	GCCTCAAGCTTTCATTACAGCGGCTTCTCACC
H_CFA_F	ATAATAACATATGCATCATCACCATCACCACGGGGAACTGAAATATAACTGTATTCTGAAT
H_CFA_R	CGTAGAAAAGCTTTTAAAGGCTCGATTCACCCCA

3.2.4 Generation of bacterial glycerol stocks

A single colony from an LB (Miller) agar plate containing transformed *E. coli* colonies was used to inoculate 5 mL LB (Miller) media in a 50 mL conical-based centrifuge tube containing an appropriate antibiotic. The tube was placed in an incubator at 37°C while shaking at 220 rpm for 16 hours. 200 of culture medium mixed with 200 50% (v/v) sterile glycerol in a sterile 1.5 mL Eppendorf tube before flash-freezing in liquid nitrogen and storing at -80°C indefinitely.

3.2.5 *E. coli* protein expression

A glycerol stock was used to inoculate a starter culture consisting of 100 mL TB (containing 50 $\mu g\,mL^{-1}$ kanamycin sulfate) in a 250 mL flask. The starter culture was incubated at 37°C for 16 hours while shaking at 225 rpm. 10 mL from the starter culture was then used to inoculate each of two 2 L flasks containing 500 mL TB. The flasks were incubated at 37°C while shaking at 225 rpm, when the OD_{600} reached 0.6 the temperature was decreased to 16°C. Protein expression was induced in both flasks when the OD_{600} reached 1 by the addition of (Isopropyl β-D-1-thiogalactopyranoside) IPTG to a final concentration of 1 mM. Following induction the flasks were left for another 16 hours before harvesting cells. Cells were harvested by centrifugation at 4000 xg and stored at -80°C until needed.

3.2.6 Purification of H_C/A1–/A3, /A5, /A6 and /FA Domains

Cell pastes were resuspended using 10 mL of Buffer A (50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole-HCl, pH 7.4) per gram of cells. Resuspended cells were passed through a cell disruptor (Constant Systems Ltd.) at 20 kPSI with the temperature maintained at 4°C. Cell debris was removed by centrifugation at 75,000 x g for 30 min at 4°C. Clarified lysate was filtered manually using 0.7 µm syringe filters (Minisart, Sartorius). Cleared lysate was loaded onto a pre-equilibrated 5 mL HisTrap HP column (GE Healthcare) at a flow rate of $5\,mL\,min^{-1}$. The column was washed with Buffer A and protein eluted using a gradient from 0% to 100% Buffer B (50 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole-HCl, pH 7.4). Fractions containing eluted protein

were pooled and concentrated using a 10 MWCO centrifugal filter (Millipore) before loading onto a Superdex200 (GE Healthcare) size-exclusion chromatography column which was pre-equilibrated with Buffer C (50 mM Tris-HCl, 500 mM NaCl, pH 7.4). The column was run at 1.5 mL min^{-1} and fractions containing H_C were pooled and concentrated using a 10 MWCO centrifugal filter (Millipore) to 1 mg mL^{-1} for storage at -20°C.

3.2.7 Purification of H_C/A4 Domain

Cell pastes were resuspended using 10 mL of Buffer A (50 mM Tris-HCl, 200 mM NaCl, 10 mM trehalose, 20 mM imidazole-HCl, pH 7.4) per gram of cells. Resuspended cells were passed through a cell disruptor (Constant Systems Ltd.) at 20 kPSI with the temperature maintained at 4°C. Cell debris was removed by centrifugation at 75,000 x g for 30 min at 4°C. Clarified lysate was filtered manually using 0.7 µm syringe filters (Minisart, Sartorius). Cleared lysate was loaded onto a pre-equilibrated 5 mL HisTrap HP column (GE Healthcare) at a flow rate of 5 mL min^{-1}. The column was washed with Buffer A and protein eluted using a gradient from 0% to 100% Buffer B (50 mM Tris-HCl, 200 mM NaCl, 10 mM trehalose, 500 mM imidazole-HCl, pH 7.4). Fractions containing eluted protein were pooled and concentrated using a 10 MWCO centrifugal filter (Millipore) before loading onto a Superdex200 (GE Healthcare) size-excluision chromatography column which was pre-equilibrated with Buffer C (50 mM Tris-HCl, 200 mM NaCl, 10 mM trehalose, pH 7.4). The column was run at 1.5 mL min^{-1} and fractions containing H_C were pooled and concentrated using a 10 MWCO centrifugal filter (Millipore) to 1 mg mL^{-1} for storage at -20°C.

3.2.8 Crystalisation of BoNT H_C Domains

Crystallisation Screening

Initial crystallisation screening for all proteins followed the following method. Crystallisation conditions were screened using commercially available 96-well screens from Molecular Dimensions, UK. The Crystal Phoenix robot (Art Robbins Instruments) was used to dispense 50 µL of each screening solution into each well of a 96-well sitting-

drop 3-drop crystallisation plate (Intelli-well®, Art Robbins Instruments). In the first, second, and third sitting drop shelf of each well was dispensed 200 nL, 200 nL, and 100 nL of screening solution respectively. To the first, second, and third sitting drop shelf of each well, 200 nL, 100 nL, and 200 nL of concentrated protein solution was dispensed to create three crystallisation ratios (1:1, 1:2, and 2:1) of protein to screening solution. The plate was sealed manually using a ClearVue™ sheet (Molecular Dimensions). Crystallisation plates were stored at 16°C.

H_C/A1

H_C/A1 was concentrated 10 mg mL^{-1} using a 10 kDa MWCO centrifugal filter (Millipore) and crystallisation conditions screened. Crystals of H_C/A1 were grown at 16°C using a 1:1 ratio of protein solution (10 mg mL^{-1}) to reservoir solution (0.1 M MIB, pH 4.0, 25% w/v PEG 1500). Crystals were soaked in cryoprotectant (equal volume of reservoir solution and 50% v/v glycerol) before vitrification in liquid nitrogen. Data were collected on beamline I04 at Diamond Light Source (UK).

H_C/A2

H_C/A2 was concentrated to 7.5 mg mL^{-1} using a 10 kDa MWCO centrifugal filter (Millipore) and crystallisation conditions were screened. Crystals were grown in the condition E9 from the SG1 screen (Molecular Dimensions, UK) (0.1 M Potassium thiocyanate, 30% w/v PEG 2000 MME). Crystals were cryoprotected using 50% v/v glycerol and reservoir solution in a 1:1 ratio before flash-cooling in liquid nitrogen. Data were collected on beamline I04-1 at Diamond Light Source (UK). 7200 images of 0.1 degree oscillation were taken with a wavelength of 0.92 Å.

H_C/A3

H_C/A3 was concentrated to 6 mg mL^{-1} using a 10 kDa MWCO centrifugal filter (Millipore) and crystallisation conditions were screened. H_C/A3 crystals were grown in condition B1 from the PACT *premier*™ screen (0.1 M MIB pH 4.0 25% w/v PEG 1500). Individual crystals were mounted in a loop and dipped in a solution containing

reservoir and 50% v/v glycerol in a 1:1 ratio for cryoprotection before flash-cooling and storage in liquid nitrogen.

H_C/A3 with GD1a

H_C/A3 was concentrated to 6 mg mL^{-1} using a 10 kDa MWCO centrifugal filter (Millipore) and GD1a (Elicityl, France) was added to a final concentration of 100 µM. The protein and ganglioside solution was used for crystallisation screening. Crystals were obtained in condition G1 (0.1 M sodium formate, 0.1 M ammonium acetate, 0.1 M sodium citrate tribasic dihydrate, 0.1 M sodium potassium tartrate tetrahydrate, 0.1 M sodium oxamate, 0.1 M imidazole, 0.1 M MES monohydrate pH 6.5, 20% v/v PEG 500 MME, 20% w/v PEG 20000) from the MORPHEUS screen (Molecular Dimensions, UK). Individual crystals were mounted in a loop and flash-cooled and stored in liquid nitrogen. Diffraction data for H_C/A3 were collected over 7200 images using 0.1° oscillations and exposures of 0.02 s on beamline IO3 at Diamond Light Source (UK).

H_C/A4

H_C/A4 was concentrated to 4 mg mL^{-1} using a 10 kDa MWCO centrifugal filter (Millipore) and crystallisation conditions screened. Crystals were grown in condition H1 (10 mM Spermine tetrahydrochloride, 10 mM Spermidine trihydrochloride, 10 mM 1,4-Diaminobutane dihydrochloride, 10 mM D-L-Ornithine monohydrochloride, 0.1 M MOPSO/bis-tris pH 6.5, 15% (w/v) PEG 3 K, 20% (v/v) 1,2,4-butanetriol, 1% (w/v) NDSB 256) from the MORPHEUS II screen Gorrec, 2015. Crystals were mounted in a loop and then flash-cooled for storage in liquid nitrogen.

H_C/A5

H_C/A5 was concentrated to 4 mg mL^{-1} using a 10 MWCO centrifugal filter (Millipore) and crystallisation conditions screened. Crystals were produced in the condition G2 (0.1 M Sodium formate, 0.1 M Ammonium acetate, 0.1 M Sodium citrate tribasic dihydrate, 0.1 M Sodium potassium tartrate tetrahydrate, 0.1 M Sodium oxamate, 0.1 M Imidazole, 0.1 M 2-(N-morpholino)ethanesulfonic acid, pH 6.5, 10% v/v Ethylene

glycol, 10% w/v PEG 8000) from the MORPHEUS screen. Crystals were flash-cooled in liquid nitrogen under a stream of nitrogen gas. Data were collected on beamline I03 at Diamond Light Source (UK). 7200 images of 0.1 degree oscillation were taken with a wavelength of 0.9763 Å.

H_C/A6

H_CA6 was concentrated to 6 mg mL^{-1} using a 10 MWCO centrifugal filter (Millipore) and crystallisation conditions screened. Crystals appeared in condition G7 (0.2 M Sodium acetate trihydrate, 0.1 M Bis-Tris propane-HCl pH 7.5, 20% w/v PEG 3350) of PACT Premier (Molecular Dimensions). Optimised crystals were produced by the increasing the precipitant concentration (0.2 M Sodium acetate trihydrate, 0.1 M Bis-Tris propane-HCl pH 7.5, 22% w/v PEG 3350) in the hanging-drop vapour diffusion method. Crystals were grown at 16°C and cryoprotected using 50% v/v glycerol and reservoir solution in a 1:1 ratio before flash-cooling in liquid nitrogen. 7200 images of 0.1 degree oscillation were taken with a wavelength of 0.9763 Å on beamline IO4 at Diamond Light Source (UK).

H_C/FA

H_C/FA was concentrated to 10 mg mL^{-1} and crystallisation conditions were screened. Crystals were grown in condition A9 (4 M sodium formate, 0.1 M sodium acetate pH 5.5) from the Clear Strategy Screen II (Molecular Dimensions). Crystals were soaked in cryoprotectant (equal volume of reservoir solution and 50% v/v glycerol) before vitrification in liquid nitrogen. 7200 images of 0.1 degree oscillation were taken with a wavelength of 0.976 25 Å on beamline IO3 at Diamond Light Source (UK).

3.2.9 X-ray Data Reduction and Structure Solution

For all proteins studied, X-ray reflections across all images were indexed and integrated using DIALS (Waterman *et al.*, 2016). Integrated data were scaled and merged using Aimless (Evans, 2006; Evans and Murshudov, 2013) from the CCP4 suite (Winn *et al.*, 2011). The phases for all X-ray diffraction data collected were solved by molecu-

lar replacement using Phaser (McCoy, 2007). Manual model building and refinement was carried out using Coot (Emsley *et al.*, 2010). Models were refined using REFMAC5 (Murshudov *et al.*, 2011) and structural validation performed using PDB_REDO (Joosten *et al.*, 2014), MolProbity (Chen *et al.*, 2010), and wwPDB Validation (Gore *et al.*, 2017). Figures of protein models were produced using either PyMOL (Schrödinger, LLC, 2015) or CCP4mg (McNicholas *et al.*, 2011).

3.3 Results and Discussion

Some of the results described in this chapter have been published within Davies *et al.* (2018c) and Davies *et al.* (2018a).

3.3.1 H_C/A1 Crystal Structure

Expression and Purification

The H_C/A1 construct was produced by sub-cloning the H_C domain DNA from a full-length BoNT/A1 sequence (UniprotKB: P0DPI1). Expression in mTB medium produced a good quantity of soluble protein - approximately 50 mg per litre of cell culture. Purification by immobilised metal affinity chromatography (IMAC) using a gradient elution resulted in a single peak of recombinant H_C/A1 (Figure 3.2a) which contained low levels of contaminants (Figure 3.3).

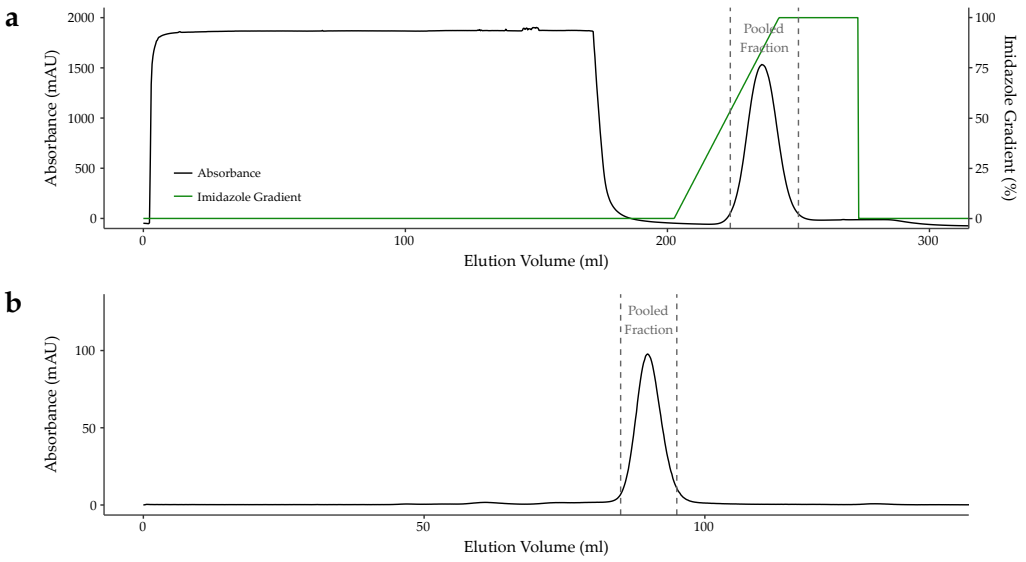


Figure 3.2: Purification of H_C/A1

(a) IMAC purification of H_C/A1 resulted in a large single peak of eluted protein. Clarified *E. coli* lysate was loaded onto a HisTrap column and the column washed before a gradient elution. **(b)** Size exclusion chromatography removed small residual contaminants and imidazole from the IMAC purification.

The whole peak was pooled and the imidazole along with residual contaminants were removed through a second stage purification using size exclusion chromatography (Figure 3.2b, Figure 3.3).

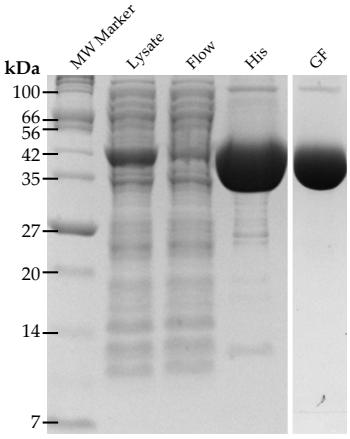


Figure 3.3: H_C/A1 Purification SDS-PAGE.
12% SDS-PAGE following purification of Hc/A1. Lane 1: Molecular weight marker. Lane 2: Crude lysate before purification. Lane 3: Unbound protein following HisTrap capture. Lane 4: Pooled fractions after His-tag purification. Lane 5: Pooled fractions after gel filtration.

Crystallisation and Data Processing

Crystallisation conditions for H_C/A1 were identified by screening the purified protein against commercially available crystallisation screens. The protein was concentrated for automated screening however at concentrations over $10\,mg\,mL^{-1}$, some of the protein auto-crystallised into small needles, visible by light microscopy. The protein was not concentrated further and used directly for crystallisation screening. Multiple hits were identified in the screen Pact *premier*^{TM} (Molecular Dimensions, UK) (Newman *et al.*, 2005). The largest of the crystals from the hits identified were found in condition B1 (0.1 M MIB, pH 4.0, 25% w/v PEG 1500). Attempts to optimise this condition through variation of pH and PEG concentration using the hanging-drop vapour diffusion method yielded poorer diffracting crystals. Instead, the condition B1 was used to fill an entire 96-well crystallisation plate in combination with H_C/A1 ($10\,mg\,mL^{-1}$) to produce further crystals for structure determination. About 30% of the wells yielded large individual crystals within one week. A second crystal morphology which appear to have rounded edges, grew within the same plate ~ six months after the initial setup (Figure 3.4).

X-ray diffraction data from both crystal morphologies were collected on beamline I04 (Diamond Light Source, UK). Initial data contained artefacts from ice-rings and so the crystals used for final data collection were cryo-protected by brief immersion in a

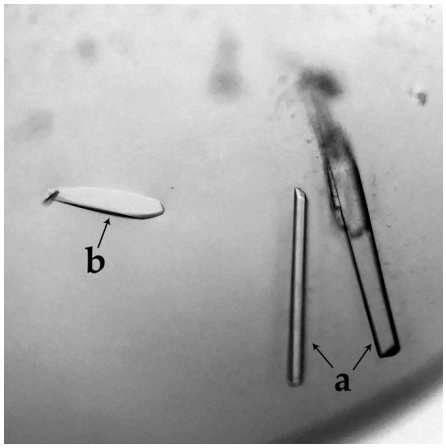


Figure 3.4: H_C/A1 Crystal Morphologies A H_C/A1 crystallisation drop from the condition 0.1 M MIB, pH 4.0, 25% w/v PEG 1500 containing two crystal morphologies. Crystal form 1 **(a)** and crystal form 2 **(b)** have the space group symmetry $P2_12_12_1$ and $P2_1$, respectively. The image shown is a composite from two photographs at different focal points from the same location.

Table 3.3: H_C/A1 Crystallographic Data Collection and Refinement Statistics

	H_C/A1 form 1	**H_C/A1** form 2
Data collection		
Beamline	I04 DLS	I04 DLS
Wavelength (Å)	0.97949	0.97949
Crystallographic statistics		
Space Group	$P2_12_12_1$	$P2_1$
Cell lengths a,b,c (Å)	39.8, 107.3, 107.6	61.4, 53.9, 62.7
Cell angles α, β, γ (°)	90.00, 90.00, 90.00	90.00, 106.10, 90.00
Resolution (Å)	24.10–1.45 (1.48–1.45)	60.24–1.80 (1.84–1.80)
R_{merge}	0.076 (0.692)	0.142 (0.539)
R_{meas}	0.097 (0.920)	0.150 (0.588)
R_{pim}	0.058 (0.601)	0.046 (0.225)
$CC_{1/2}$	0.997 (0.357)	0.962 (0.954)
Mean $< I/\sigma I >$	6.9 (1.3)	14.4 (3.7)
Completeness (%)	95.3 (95.1)	98.6 (89.1)
Total reflections	261047	739637
Total unique reflections	78738	36131
Multiplicity	3.3 (2.7)	20.4 (11.7)
Refinement statistics		
R_{work}/R_{free}	0.176/0.221	0.188/0.223
RMSD bond lengths (Å)	0.002	0.005
RMSD bond angles (°)	0.44	0.73
Ramachandran statistics (%)		
Favoured	96.3	96.4
Allowed	3.5	3.6
Outliers	0.2	0.0
Wilson B-factors (Å^2)	15.9	13.4
Average B-factors (Å^2)		
Protein	21.2	22.5
Water	34.1	28.0
No. of atoms		
Protein	3945	3299
Water	421	322
PDB Code	5MK6	5MK7

solution containing 50% v/v condition B1 and 50% v/v glycerol before cryo-cooling in liquid N_2. Complete datasets for both crystal morphologies, termed crystal form 1 and 2, were collected after cryo-protection and the best of each crystal form diffracted to 1.45Å and 1.8Å, respectively. The data was indexed and integrated using DIALS (Winter *et al.*, 2018); the CCP4 program POINTLESS (Evans, 2006) was used to identify the correct space groups. The two crystal morphologies were of different space groups, with crystal form 1 determined to be in the orthorhombic $P2_12_12_1$ space group while crystal form 2 was of the $P2_1$ space group (Table 3.3).

Structure Refinement and Analysis

The phases for initial structural refinement of both datasets were produced by molecular replacement using an existing H_C/A1 crystal structure (PDB ID: 2VUA, Stenmark *et al.*, 2008). Both structures are generally very similar and superpose well with an RMSD between all atoms of 0.78 Å (Figure 3.5). There are some striking differences between the two crystal forms, the first of these is at the N-terminus (Figure 3.5b). H_C/A1 was crystallised with a 6x histidine-tag present at the N-terminus but no density is present for either crystal form likely due to disorder within the crystals. For crystal form 1, the first residue is not visible but the following residues form a small loop before joining the jelly-roll fold, as previously identified in BoNT/A1 (Lacy *et al.*, 1998) which consists of fourteen β-strands. For crystal form 2 however, the first fourteen residues after the histidine-tag have no clear density and residues 889–894 form a small α helix. As a result the typical jelly-roll fold is missing two of the fourteen β-strands. The significance of this is not known but it is noteworthy that the translocation of BoNTs is not well understood and the N-terminal residues of the H_C may play a role in translocation with the H_N domain.

Immediately obvious in the resulting electron density maps for the crystal form 2 was a backbone rearrangement, compared with crystal form 1, in the region near to the GBS which allowed two nearby cysteine residues (Cys_{1235} and Cys_{1280}) to form a covalent disulfide bond (Figure 3.6). This has also been observed in the full length crystal structure of BoNT/A1 (PDB ID: 3BTA, Lacy *et al.*, 1998). The density of the sulfur atom from Cys_{1235} is weaker than that of Cys_{1280} but this may be due to X-ray induced radiation damage of the disulfide bond which is generally more sensitive

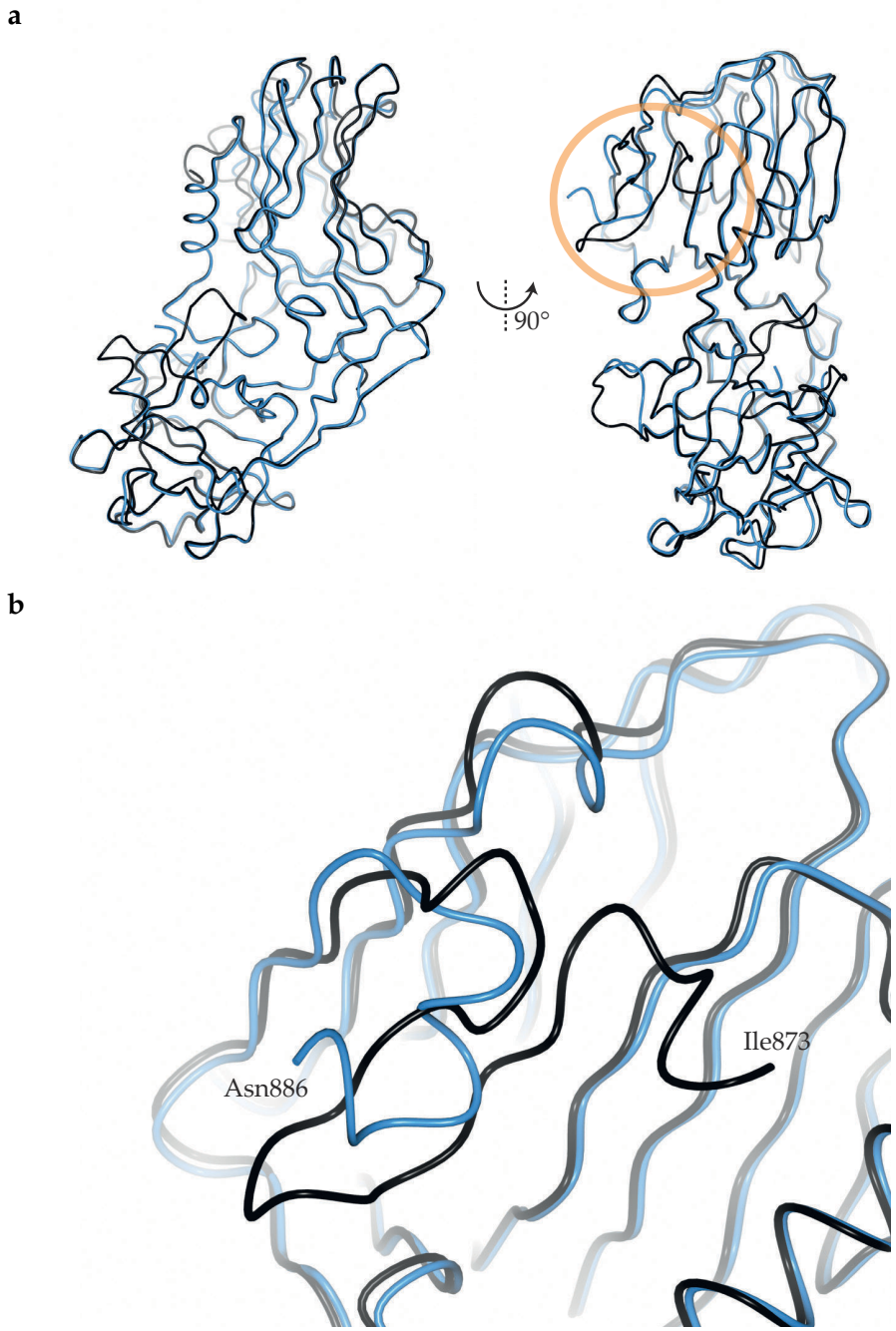


Figure 3.5: H_C/A1 Crystal Structures

(a) The crystal structures from both crystal form 1 (black) and crystal form 2 (blue) of H_C/A1 superposed to display similarity. **(b)** The N-terminus of both structures (Asn886 and Ile873) diverge significantly.

than other protein atoms (Petrova *et al.*, 2010). No disulfide bond was observed in the structure from the crystal form 1. Instead, Cys_{1235} rotated away from the Cys_{1280} due to a backbone rotation, bringing the neighbouring Lys_{1236} toward Cys_{1280} (Figure 3.6). This is consistent with previous H_C/A1 structures either in complex with GT1b (PDB ID: 2VU9, Stenmark *et al.*, 2008), SV2-LD (PDB ID: 4JRA, Benoit *et al.*, 2014; 5JLV, Yao *et al.*, 2016) or in the apo form (PDB ID: 2VUA; Stenmark *et al.*, 2008). This finding suggests that the crystallisation condition is not the only determinant as to whether or not the disulfide bond in the H_C is formed. The two cysteine residues in question are highly conserved across all BoNT subtypes which suggests they are of high importance, however their function is unknown.

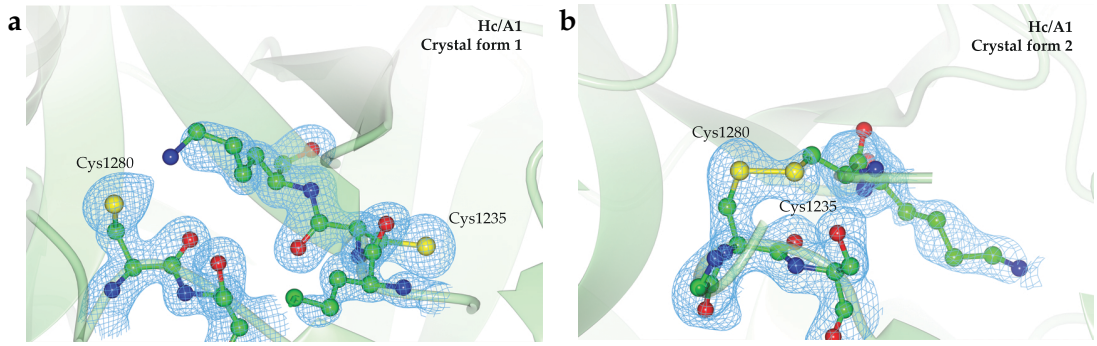


Figure 3.6: Location of Cys1235 and Cys1280 residues in H_C/A1
Cys1235 and Cys1280 are not covalently bound in the H_C/A1 crystal form 1 **(a)**. A disulphide bond is formed in crystal form 2 **(b)**. Composite OMIT maps were produced using Phenix (Terwilliger *et al.*, 2008)) and are shown for each model at 1σ. Figure generated using CCP4mg (McNicholas *et al.*, 2011). Figure reproduced from Davies *et al.* (2018a).

They may be required for disulfide bond formation to aid structural stability or for efficient function of a nearby ganglioside-binding pocket. It is important to note that the previous crystal structure determined of H_C/A1 in complex with a ganglioside analogue of GT1a does not contain the disulfide bond and therefore binding does not require the bond to be present (Stenmark *et al.*, 2008). In addition, the protein expression system does not appear to select for reduced or non-reduced forms of these cysteine residues – each have been observed in different structures of BoNT/A purified from both the native *C. botulinum* and *E. coli* (Garcia-Rodriguez *et al.*, 2007; Stenmark *et al.*, 2008; Fu *et al.*, 2009; Gu *et al.*, 2012; Przedpelski *et al.*, 2013; Benoit

et al., 2014; Yao *et al.*, 2017; Davies *et al.*, 2018a).

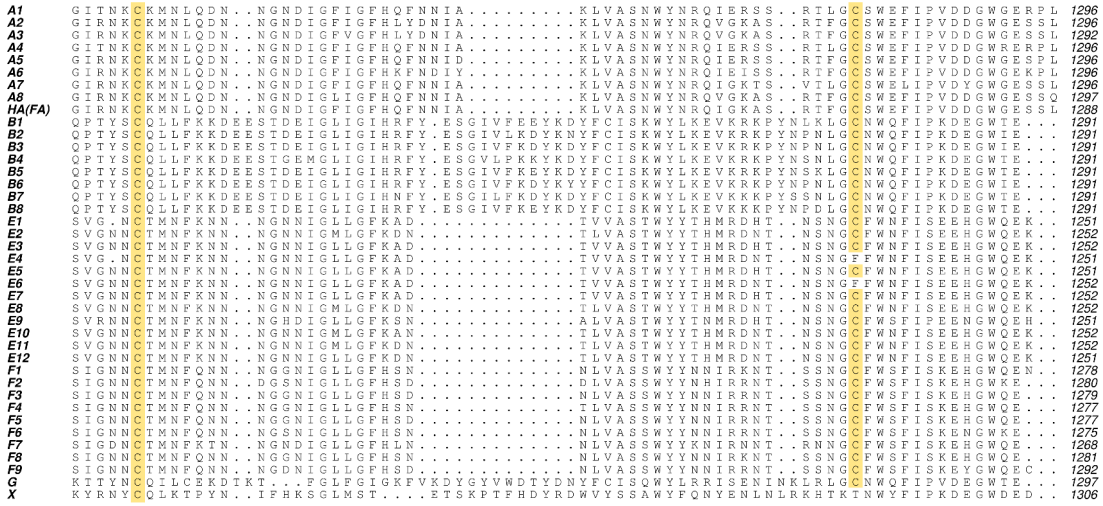


Figure 3.7: Conservation of Cysteines in BoNT H_C Domains

Amino acid sequence alignment of BoNT subtypes emphasising the strong conservation of two cysteine residues near the protein C-terminus. UniprotKB accession numbers – A1: A5HZZ9, A2: Q45894, A3: Q3LRX9, A4: Q3LRX8, A5: C1IPK2, A6: C9WWY7, A7: K4LN57, A8: A0A0A7PDB7, HA(FA): WP_047402807*, B1: P10844, B2: A0A0B4W2B0, B3: A2I2S2, B4: A2I2S4, B5: A0A0E1L271, B6: H3K0G8, B7: H9CNK9, B8: M9VUL2, E1: Q00496, E2: A2I2S6, E3: A0A076L133, E4: P30995, E5: Q9K395, E6: A8Y878, E7: G8I2N7, E8: G8I2N8, E9: WP_017352936*, E10: A0A076JVL9, E11: A0A076K0B0, E12: A0A0A7RCR1, F1: A7GBG3, F2: Q9ZAJ5, F3: D2KHR6, F4: D2KHQ8, F5: D2KHQ9, F6: D2KHS6, F7: D2KHS9, F8: WP_-076177537 *, F9: A0A1P8YWK9, G: Q60393, X: WP_045538952* (*indicates NCBI accession code where UniprotKB accession is not available). Figure reproduced from Davies *et al.* (2018a).

It is unusual, but not unheard of, for the same protein in the same condition to produce multiple crystal forms of different space groups. The second crystal form may have appeared over time as a result of a slight change to the crystallisation condition as its components degrade. However, this is unlikely and in fact the second crystal form was found to differ from the first by a single disulfide bond within the structure and it may simply have been present at a lower concentration, requiring a longer period to crystallise.

3.3.2 H_C/A2 Crystal Structure

The H_C/A2 construct was produced by sub-cloning the H_C domain DNA from a full-length BoNT/A2 sequence (UniprotKB: Q45894). Expression produced a low quantity of soluble protein when compared to H_C/A1 but still plenty for structural biology experiments. Purification by immobilised metal affinity chromatography (IMAC) using a gradient elution resulted in a small single peak of recombinant H_C/A2 (Figure 3.8a). The protein peak was pooled before desalting and further purification using size exclusion chromatography (Figure 3.8b).

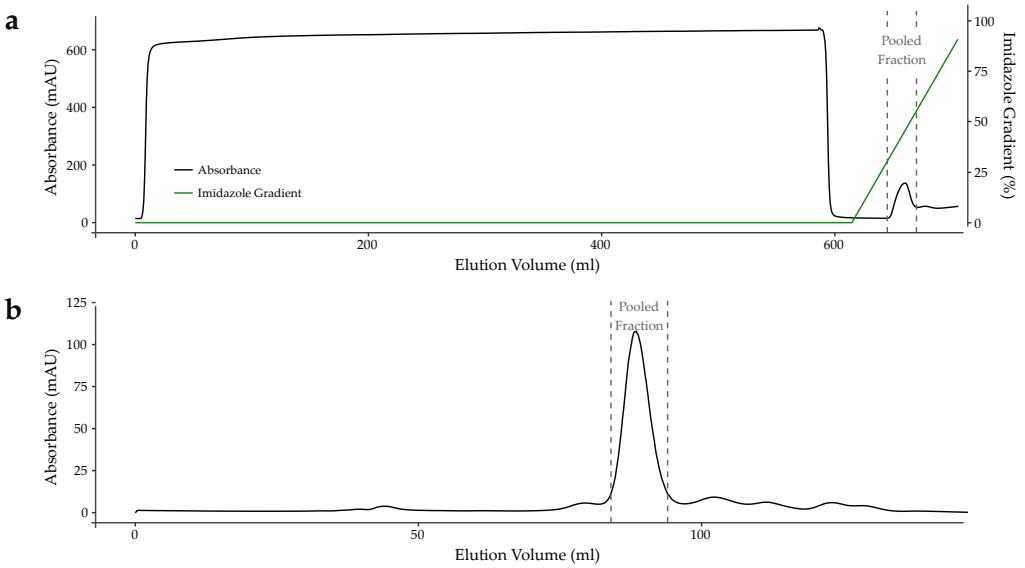


Figure 3.8: Purification of H_C/A2

(a) IMAC purification of H_C/A2 resulted in a single peak of eluted protein. Clarified *E. coli* lysate was loaded onto a HisTrap column and the column washed before a gradient elution. **(b)** Size exclusion chromatography removed small residual contaminants and imidazole from the IMAC purificaiton.

Crystallisation and Data Processing

Crystallisation conditions for H_C/A2 were found by screening the purified protein against commercially available crystallisation screens. A single hit was identified in the SG1 screen (Molecular Dimensions, UK) (Newman *et al.*, 2005). Crystals appeared in the condition E9 (0.1 M Potassium thiocyanate, 30% w/v PEG 2000 MME) after sixteen

months. Attempts to optimise this condition in 24-well hanging-drop crystallisation plates containing a varied salt and PEG concentration failed to produce crystals. Data from multiple crystals were collected on beamline I04-1 of Diamond Light Source. Crystals were of space group $P2_12_12_1$ and the best produced good diffraction data to 1.75 Å. A total of 9999 frames, of low-dose fine-sliced data (0.1° oscillation) were collected. The data exhibited some artefacts which were likely due to low levels of ice-rings in the recorded diffraction data despite the addition of a cryoprotectant.

Structure Refinement and Analysis

A model was produced from the H_C/A2 protein sequence using Phyre2 web server (Kelley *et al.*, 2015) for molecular replacement. H_C/A2 shares 87.29% identity with H_C/A1 (Table 3.1) with some of those residue differences covering the SV2 binding-site and the SV2-glycan binding-site (Figure 3.1). The X-ray structure was solved and with a single protein molecule in the asymmetric unit. The overall fold of H_C/A2 is highly similar to H_C/A1 with an RMSD between C_α atoms of 0.46 Å. The electron density was generally very good with the exception of three small loop regions which could not be modelled (Val_{1271}–Arg_{1276}, Asp_{1228}–Ile_{1231}, Tyr_{1165}–Asp_{1170}). This is the highest resolution H_C/A2 structure determined to date; two other crystal structures of H_C/A2 in complex with non-glycosylated SV2 have previously been determined to 2.3 Å (PDB ID: 5MOY, Benoit *et al.*, 2017) and 2.0 Å (PDB ID: 6ES1, Gustafsson *et al.*, 2018). The structure of H_C/A2 in complex with non-glycosylated SV2 determined by Gustafsson *et al.* (2018) suggests the SV2 binding-site may have some flexibility and is able to accommodate small angular deviations in binding. Our structure is the first of H_C/A2 without its protein receptor bound and displays little structural change from the previously structures. Any flexibility in the SV2-binding region is likely due to crystal packing and/or whether the SV2 peptide is correctly glycosylated. Our structure also contained a single disulfide bond between Cys_{1235} and Cys_{1280}. Interestingly, of the two previous H_C/A2 structures, only one (PDB ID: 6MOY) has the same disulfide bond present.

Table 3.4: H_C/A2 Crystallographic Data Collection and Refinement Statistics

	H_C/A2
Data collection	
Beamline	I04-1 DLS
Wavelength (Å)	0.91587
Crystallographic statistics	
Space Group	$P2_12_12_1$
Cell lengths a,b,c (Å)	39.71, 99.81, 115.54
Cell angles α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	75.53–1.75 (1.78–1.75)
R_{merge}	0.187 (1.274)
R_{meas}	0.193 (1.323)
R_{pim}	0.048 (0.353)
$CC_{1/2}$	0.999 (0.430)
Mean $< I/\sigma I >$	13.6 (3.0)
Completeness (%)	100 (99.8)
Total reflections	1,408,426 (67,617)
Total unique reflections	47,356 (2,538)
Multiplicity	29.7 (26.6)
Refinement statistics	
R_{work}/R_{free}	0.203/0.255
RMSD bond lengths (Å)	0.02
RMSD bond angles (°)	1.95
Ramachandran statistics (%)	
Favoured	96.30
Allowed	3.45
Outliers	0.25
Wilson B-factors (Å^2)	20.8
Average B-factors (Å^2)	
Protein	27.25
Water	40.282
No. of atoms	
Protein	3360
Water	153
PDB Code	–

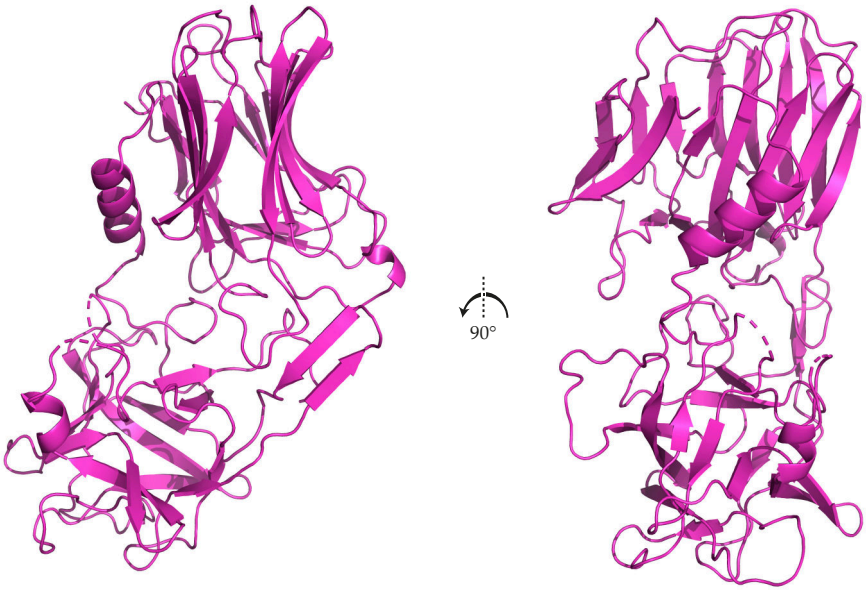


Figure 3.9: H_C/A2 Crystal Structure

Cartoon representation of the H_C/A2 crystal structure. Missing loops are displayed as dashed lines.

3.3.3 H_C/A3 Crystal Structure

Protein Purification

No structural information for H_C/A3, which shares 86.82% identity with H_C/A1 (Table 3.1), was available and so we attempted crystallisation of the protein in order to determine whether any sequence differences in BoNT/A3 would affect the structure of either the ganglioside or SV2 binding-sites. The H_C/A3 construct was produced by sub-cloning the H_C domain DNA from a full-length BoNT/A3 sequence (UniprotKB: Q3LRX9) and adding an N-terminal histidine tag. Expression produced a good quantity of soluble protein which was first purified IMAC using a gradient elution resulting in multiple but well separated protein peaks (Figure 3.10a). The protein peak corresponding to H_C/A3 was pooled before and further purified using size exclusion chromatography (Figure 3.10b).

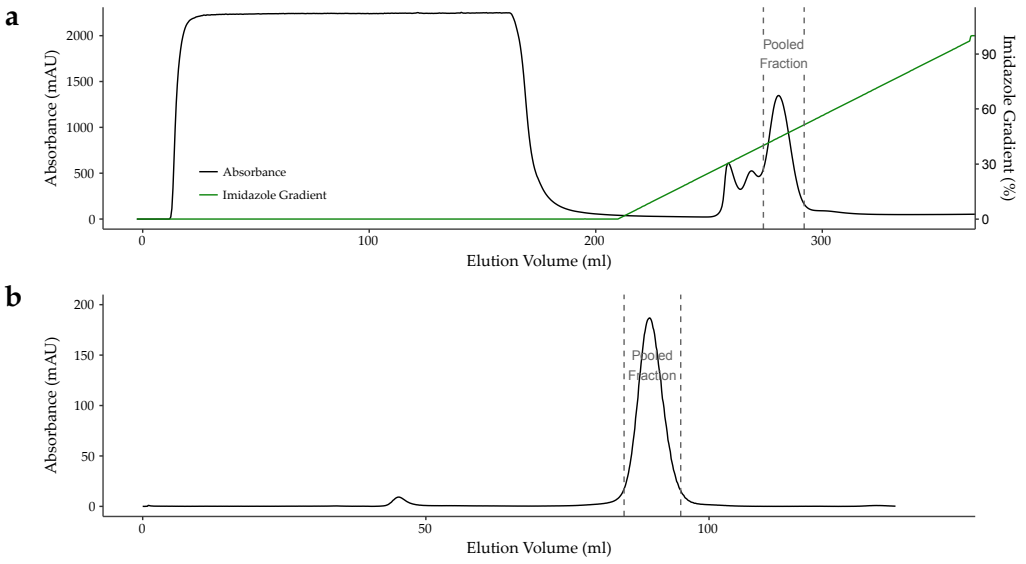


Figure 3.10: Purification of H_C/A3

(a) IMAC purification of H_C/A3 resulted in multiple peaks of eluted protein. Clarified *E. coli* lysate was loaded onto a HisTrap column and the column washed before a gradient elution. **(b)** Size exclusion chromatography removed residual contaminants and imidazole from the IMAC purification resulting in a clean single peak.

Crystallisation and Data Processing

The protein was concentrated for crystallisation, some precipitation was present at concentrations above 7 mg mL^{-1}, therefore crystallisation trials were attempted at a concentration of 6 mg mL^{-1}. After one week, some small crystals appeared in the condition B1 (0.1 M MIB pH 4.0, 25% w/v PEG 1500) from the Pact *premier*^{TM} screen (Molecular Dimensions, UK). The crystals grew in size significantly and resembled long, individual rods (Figure 3.11). Interestingly this is the same condition that produces good quality crystals of H_C/A1 (Figure 3.4), with both H_C/A1 crystal form 1 and H_C/A3 sharing the same crystal morphology. Three other crystallisation conditions were also later identified which produced crystals of a similar appearance.

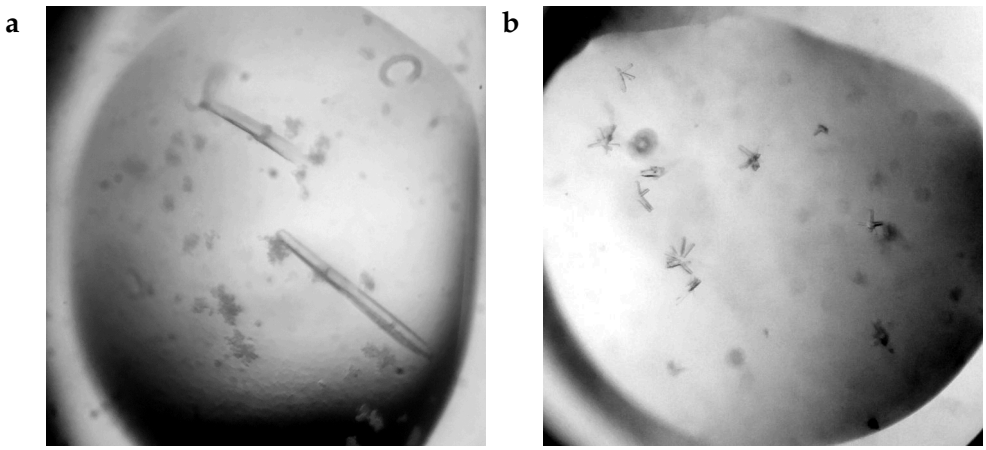


Figure 3.11: H_C/A3 Crystal Morphologies
(a) Crystals of H_C/A3 grew into long individual rods. **(b)** Crystalisation of the H_C/A3–GD1a analogue complex produced very small clusters of rods which were possible to manually break into invidividual crystals.

X-ray diffraction data were collected from multiple crystals, with the best data arising from crystals that were cryoprotected using glycerol in the same manner as H_C/A1 crystals. In total, 11 datasets were collected from 11 individual crystals obtained across multiple conditions with the best producing usable data to 1.6 Å (Table 3.6). The best dataset was manually processed to determine a resolution cut-off. The data were indexed and integrated using DIALS into the space group $P2_12_12_1$. Data were scaled and merged using AIMLESS and a resolution cut-off of 1.6 Å was imposed based on all available statistics with particular attention to the outer-shell $CC_{1/2}$ and

Table 3.5: X-ray Datasets Collected for H_C/A3

Screen (Well)	Composition	Crystals (Datasets)	Auto Resolution
PP† (B1)		1 (3)	1.54 Å
PP† (B10)		1 (0)	– ‡
MORP† (B9)		7 (5)	1.55 Å
SG1 (H5)		2 (2)	1.84 Å

† Screen names are abbreviated. PP: Pact premier; MORP: Morpheus. ‡ Autoprocessing failed to complete.

completeness (Evans and Murshudov, 2013). To produce the electron density map, the data were solved by molecular replacement using the Phaser software (McCoy *et al.*, 2007). A starting model was produced by homology modelling using the Phyre2 web server (Kelley *et al.*, 2015) – the existing H_C/A1 structure (PDB ID: 2VUA) was used as a template for the H_C/A3 sequence.

Structure Refinement and Analysis

The electron density map following molecular replacement was very clear for most of the structure with exception of three loop regions between residues: 1198–1201, 1268–1277 and 1227–1231. It is likely that disorder is the cause as none of these regions are involved in crystal packing. The H_C/A3 protein sequence is most similar to that of the BoNT/A2 subtype, where both binding domains share 98.8% identity. This is also reflected by a high structural similarity between the two, where the RMSD between Cα atoms is 0.83 Å.

Recent structures of both H_C/A1 and H_C/A2 bound to their protein receptor, SV2 isoform C (SV2C), suggest that the overall binding mode is generally conserved (Benoit *et al.*, 2014; Yao *et al.*, 2016; Benoit *et al.*, 2017). The high conformational similarity with our structure suggests that H_C/A3 may also bind SV2C in the same manner. One notable point of variation in sequence however, is position 1156 (BoNT/A1 numbering, Figure 3.1). BoNT/A3 (and subtypes /A4 to /A8) contain a methionine which has the potential to form an electrophilic interaction or weak hydrogen bond with H563 of SV2C (Pal and Chakrabarti, 2001), whereas BoNT/A1 has an arginine that forms a cation-π stacking interaction, and BoNT/A2 has a glutamate that forms a salt bridge with the equivalent histidine of SV2C (Benoit *et al.*, 2014; Yao *et al.*, 2016; Benoit

Table 3.6: H_C/A3 Crystallographic Data Collection and Refinement Statistics

	H_C/A3	**H_C/A3 – GD1a**
Data collection		
Beamline	I03 DLS	I04 DLS
Wavelength (Å)	0.97622	0.9795
Crystallographic statistics		
Space Group	$P2_12_12_1$	$P2_1$
Cell lengths a,b,c (Å)	39.76, 96.47, 110.98	45.56, 72.47, 144.74
Cell angles α, β, γ (°)	90.00, 90.00, 90.00	90.00, 94.82, 90.00
Resolution (Å)	96.47–1.60 (1.63–1.60)	72.31–2.15 (2.21–2.15)
R_{merge}	0.126 (1.352)	0.313 (1.924)
R_{meas}	0.131 (1.415)	0.366 (2.286)
R_{pim}	0.036 (0.408)	0.187 (1.214)
$CC_{1/2}$	0.997 (0.639)	0.985 (0.415)
Mean $< I/\sigma I >$	10.3 (1.9)	4.9 (1.0)
Completeness (%)	96.9 (79.6)	99.8 (98.2)
Total reflections	715,684 (25,075)	372,245 (27,398)
Total unique reflections	55,445 (2,169)	51,328 (4,114)
Multiplicity	12.9 (11.6)	7.3 (6.7)
Refinement statistics		
R_{work}/R_{free}	0.179/0.214	0.243/0.265
RMSD bond lengths (Å)	0.016	0.01
RMSD bond angles (°)	1.819	1.36
Ramachandran statistics (%)		
Favoured	97.74	95.46
Allowed	2.26	4.41
Outliers	0	0.13
Average B-factors (Å^2)		
Protein	23.5	29.25
GD1a	–	79.47
Water	33.8	23.50
No. of atoms		
Protein	3319	6700
Water	364	195
PDB Code	6F0O	–

et al., 2017). This substitution may have an effect on the binding affinity of SV2 but a structure of the SV2 complex is needed to accurately describe any potential interaction. N-linked glycosylation of SV2C is also important for binding of BoNT/A1 to the neuronal cell membrane (Yao *et al.*, 2016). Under normal physiological conditions SV2 is glycosylated. The glycan forms an array of interactions with BoNT residues that neighbour those which interact with the protein chain directly. Within BoNT/A3 (and subtypes /A2, A6 and /A8), a mutation to a residue involved in glycan binding, H1064R (BoNT/A1 numbering, Figure 3.1), may cause a change to glycan binding specificity, or indicate a potentially different but conserved binding mode to a glycan compared to BoNT/A1.

The ganglioside binding site of BoNT/A1 appears to be mostly conserved in the H_C/A3 structure except for a tyrosine at position 1113 (instead of phenylalanine) and an unresolved conformation of R1276 – which could not be confirmed due to poor electron density. Considering that BoNT/A1 binds to GT1b (Stenmark *et al.*, 2008), it is highly likely that BoNT/A3 binds gangliosides in a similar manner however the tyrosine to phenylalanine substitution would likely result in the loss of hydrogen bonding to the Sia5 moiety.

To further analyse ganglioside binding to H_C/A3 we attempted to produce a crystal structure of the protein in complex with a ganglioside analogue. Crystallisation screens were setup in which H_C was mixed with an equimolar quantity of a GD1a analogue which lacked the ceramide group (Elicityl, France). We were successful in attaining a crystallisation condition for H_C/A3 in complex with the GD1a analogue. The sugar part of GD1a differs from GT1b by a single sialic acid, Sia7, which is not involved in the BoNT/A1 to ganglioside interaction (Stenmark *et al.*, 2008). Crystals produced for the complex were extremely small and diffracted to a resolution of 2.15 Å (Figure 3.11). Unlike the H_C/A3 crystal, the complex was in the space group $P2_1$ and contained two molecules in the asymmetric unit. Following molecular replacement using the H_C/A3 structure, some unmodelled density around the expected ganglioside binding site was observed indicating bound ligand, however modelling of any part of a ganglioside was challenging and precise atom locations were not immediately obvious. While still weak, the electron density did improve after multiple rounds of refinement and most of the GD1a molecule could be placed with exception to Sia6 for which there was no apparent density (Figure 3.13).

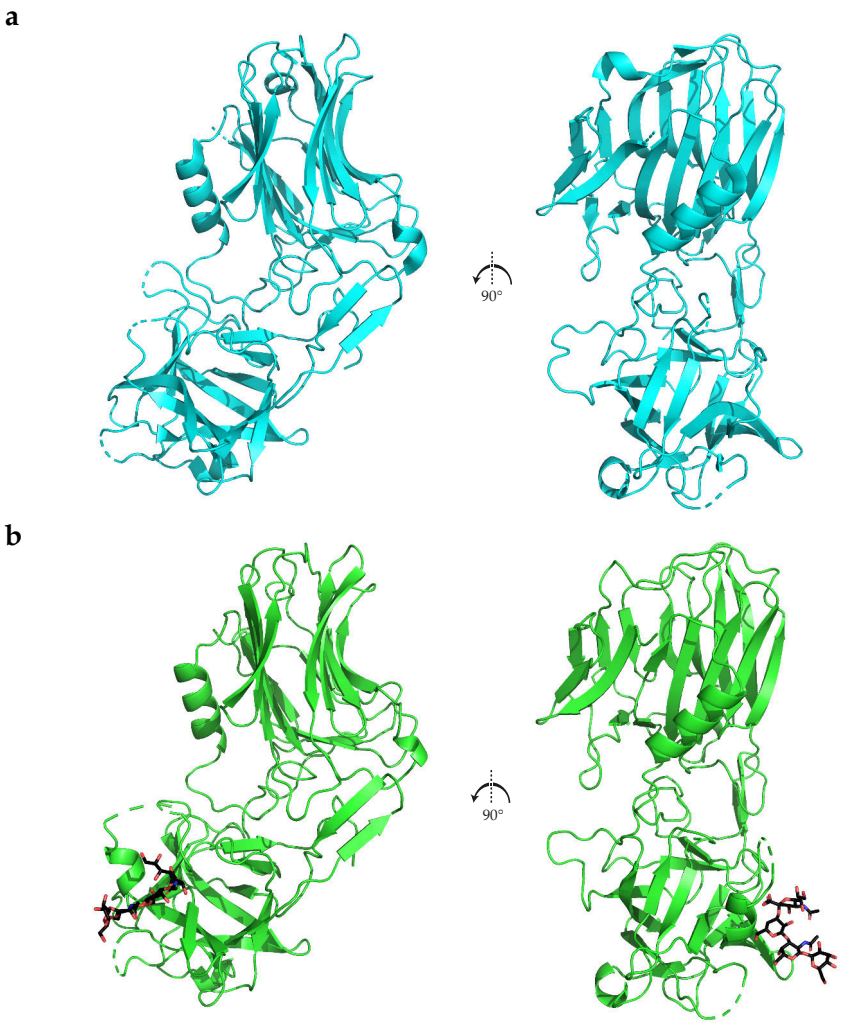
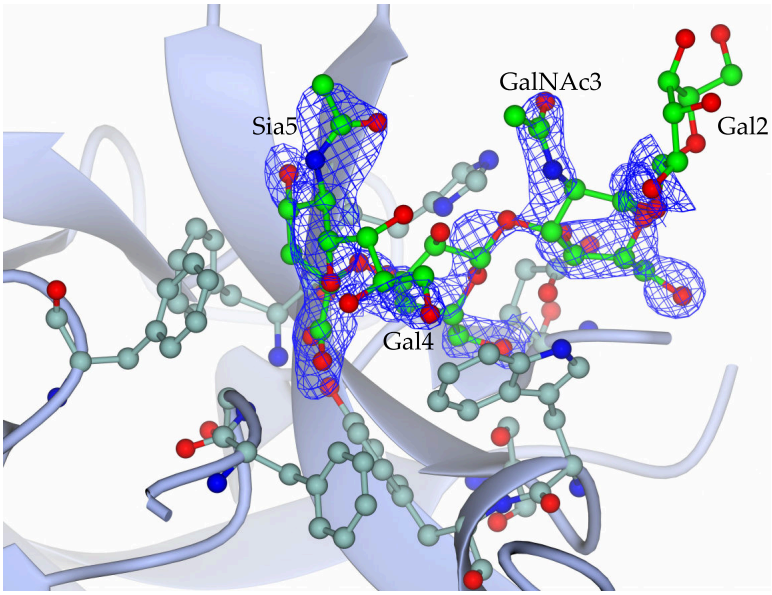


Figure 3.12: H_C/A3 Crystal Structures

Overall cartoon representation of two crystal structures: **(a)** H_C/A3 (PDB ID: 6F0O, Davies *et al.*, 2018c) and **(b)** H_C/A3 in complex with GD1a. GD1a displayed as black sticks.

a



b

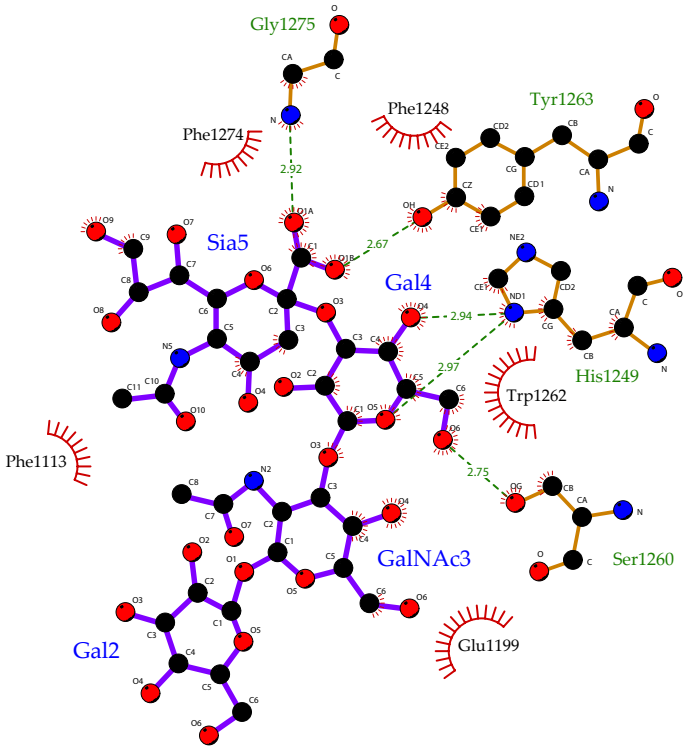


Figure 3.13: H_C/A3–GD1a Electron Density and Interactions

(a) GD1a (green sticks) located in the ganglioside binding site of H_C/A3 (blue ribbon and sticks). 2mFo–Fc electron density is shown for GD1a at 1σ. **(b)** Interactions of the GD1a ganglioside with H_C/A3. Hydrogen bonds are illustrated with green dashed lines. Hydrophobic interactions are displayed as red crescents.

Of the modelled GD1a atoms, hydrogen bond interactions are only visible for Sia5 and Gal4 of the GD1a molecule. The Phe at position 1113 has maintained a similar orientation to the equivalent Tyr in BoNT/A1 but now only contributes hydrophobic interactions. The loss of hydrogen bonding to Sia5 when compared to H_C/A1 does not seem to have had an impact structurally on the ganglioside binding.

The crystal structure of H_C/A3 was published (Davies *et al.*, 2018c) and deposited in the PDB under the accession code 6F0O.

3.3.4 H_C/A4 Crystal Structure

Protein Purification

Initial attempts to purify H_C/A4 using the same conditions as other H_C/A subtypes resulted in low levels of soluble protein due to protein precipitation. Expression levels were generally lower than other H_C/A subtypes and optimisation of the purification was required. It was determined that H_C/A4 was more stable in a decreased salt buffer that contained the additive, trehalose. Under these conditions a larger quantity of protein was produced which showed increased solubility (Figure 3.14).

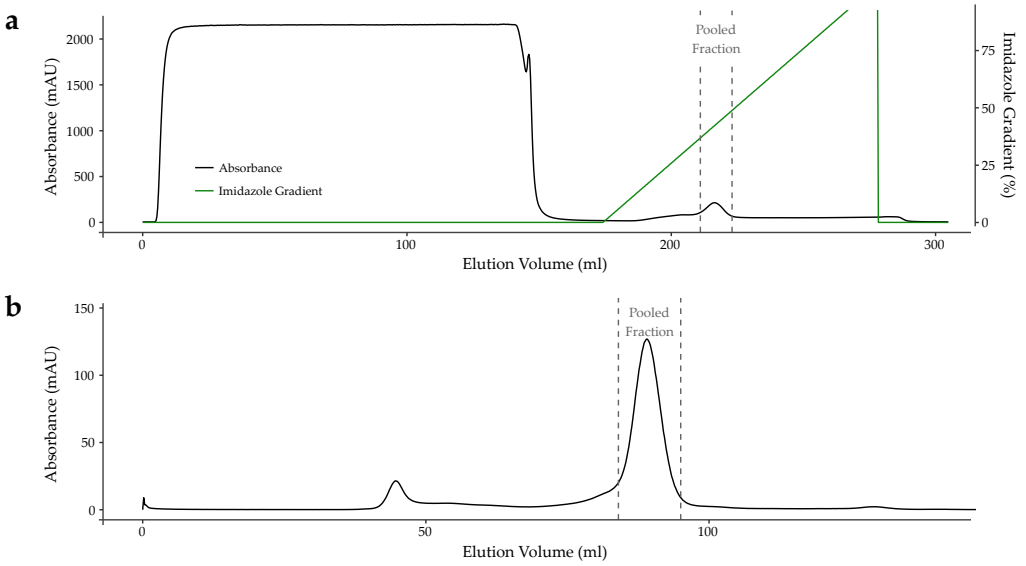


Figure 3.14: Purification of H_C/A4

(a) IMAC purification of H_C/A4 resulted in a small peak of eluted protein. Clarified *E. coli* lysate was loaded onto a HisTrap column and the column washed before a gradient elution. **(b)** Size exclusion chromatography was performed to removed small residual contaminants and imidazole from the IMAC purification.

Crystallisation and Data Collection

H_C/A4 was able to be concentrated to 4 mg mL^{-1} before signs of protein precipitation were present. Crystallisation conditions were screened and multiple hits were identified. The best crystal of approximately 60 µm^3 diffracted to 1.34 Å (Figure 3.15).

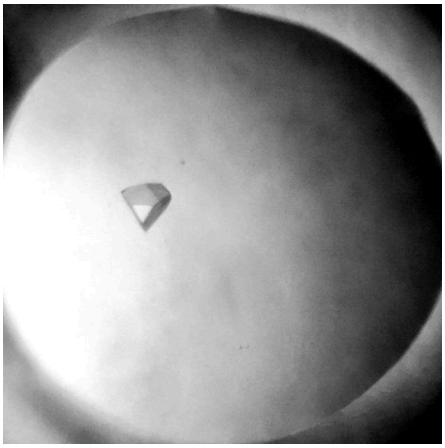


Figure 3.15: H_C/A4 Crystal Morphology
Crystallisation of H_C/A4 produced a single well defined crystal.

Structure Refinement and Analysis

The crystal belonged to space group $P2_13$ with one molecule in the asymmetric unit (Table 3.7) (Figure 3.16). A model of H_C/A4 was produced using the Phyre2 web server (Kelley *et al.*, 2015) which was used to solve the structure by molecular replacement. The exceptional quality of the electron density map allowed modelling of all residues, as well as the clear identification of a single disulfide bond between C_{1235} and C_{1280}. Of all the different subtypes of BoNT/A, the binding domain of BoNT/A4 is most similar to that of BoNT/A1 sharing 91.78% sequence identity and a RMSD of 0.72 Å between Cα atoms.

After multiple rounds of model building a large region of unmodelled electron density remained which appeared attached to His_{1259}. The density was eventually modelled as a nickel atom coordinated by the N_δ of His_{1259} and a single bis-Tris methane (Bis[2-hydroxyethyl]amino-tris[hydroxymethyl]methane) molecule (Figure 3.17). Bis-Tris methane is a common buffer which is capable of complexation of metals including nickel (Ferreira *et al.*, 2015). The metal coordination was validated using the CheckMyMetal server (Center for Structural Genomics of Infectious Diseases, US) which confirmed octahedral geometry and found the bond distances to fall within accepted ranges (Zheng *et al.*, 2017). The source of the metal was likely from the IMAC column which contains immobilised nickel, while the bis-tris methane buffer was present in the crystallisation condition. Interestingly, while this feature is not biologically

Table 3.7: H_C/A4 Crystallographic Data Collection and Refinement Statistics

	H_C/A4
Data collection	
Beamline	I03 DLS
Wavelength (Å)	0.97628
Crystallographic statistics	
Space Group	$P2_13$
Cell lengths a,b,c (Å)	116.00, 116.00, 116.00
Cell angles α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	81.95–1.34 (1.36–1.34)
R_{merge}	0.091 (2.306)
R_{meas}	0.093 (2.424)
R_{pim}	0.020 (0.740)
$CC_{1/2}$	0.999 (0.421)
Mean $< I/\sigma I >$	17.5 (1.1)
Completeness (%)	100 (100)
Total reflections	2,516,326 (60,014)
Total unique reflections	116,165 (5,700)
Multiplicity	21.7 (10.5)
Refinement statistics	
R_{work}/R_{free}	0.144/0.158
RMSD bond lengths (Å)	0.001
RMSD bond angles (°)	1.490
Ramachandran statistics (%)	
Favoured	96.62
Allowed	3.38
Outliers	0
Average B-factors (Å^2)	
Protein	20.7
Water	30.8
No. of atoms	
Protein	3504
Water	309
PDB Code	6F0P

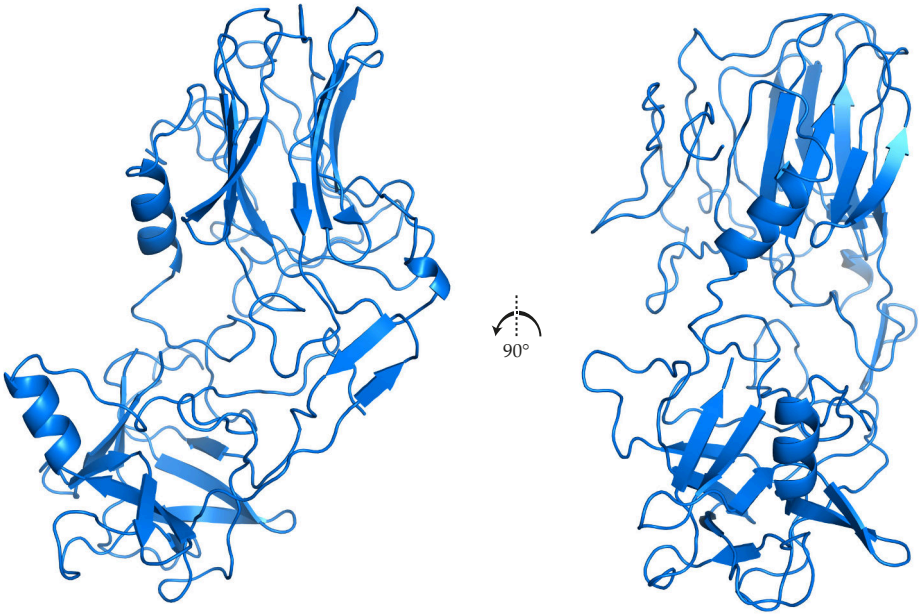


Figure 3.16: H_C/A4 Crystal Structure

Overall cartoon representation of H_C/A4.

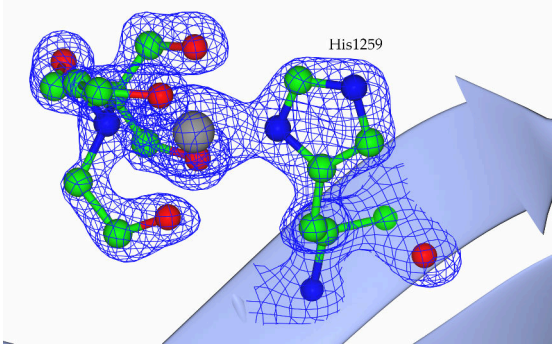


Figure 3.17: H_C/A4 bis-Tris

Electron density contoured to 1σ for a Ni atom coordinated by His_{1259} and a single molecule of bis-tris methane.

relevant, it is extremely rare and only five other structures in the PDB have a modelled bis-Tris methane, nickel, and histidine coordination complex. All of the equivalent residues involved in ganglioside binding for BoNT/A1 (Stenmark *et al.*, 2008) remain unchanged in H_C/A4 suggesting the BoNT/A4 may share the same ganglioside binding preference as BoNT/A1.

When comparing binding sites, BoNT/A4 is predicted to have the both the same ganglioside preference as BoNT/A1 and is also predicted to bind to SV2C, according to sequence alignment (Figure 3.1). Most interactions to the SV2C peptide are likely unchanged owing to the nature of backbone-backbone interactions between β-strands, however like H_C/A3, the H_C/A4 structure shows a methionine at position 1156 rather than an arginine at the equivalent position for H_C/A1, which may form an important interaction with SV2C. This may have an important impact on binding as Arg_{1156} of H_C/A1 forms electron stacking interactions with an aromatic side chain of SV2.

3.3.5 H_C/A5 Crystal Structure

Protein Purification

We attempted crystallisation of H_C/A5 which shares 93.9% identity with H_C/A1 (Table 3.1). A construct was produced by subcloning the H_C domain from the full-length BoNT/A5 sequence (UniprotKB: C7BEA8) and adding an N-terminal His tag. Expression produced soluble protein which was purified by IMAC using a gradient elution resulting in a peak corresponding to H_C/A5 (Figure 3.18a). This peak was pooled and further purified using size exclusion chromatography (Figure 3.18b) to give a highly pure product (Figure 3.19).

Crystallisation and Data Processing

H_C/A5 was concentrated to 4 mg mL^{-1} for crystallisation experiments. Some precipitation was visible for concentrations higher than this. Crystals initially appeared in the condition G5 of The MORPHEUS screen (Molecular Dimensions) and manual optimisation was attempted. After another month, further crystallisation hits were identified and X-ray diffraction data collection was attempted for each. Crystals from condition G5 of The MORPHEUS screen diffracted best to 1.14Å. Crystals from the original hit and optimisation screen all produced highly anisotropic data for which it was not possible to find an indexing solution. The structure was solved by molecular replacement using a model produced on the Phyre2 web server (Kelley *et al.*, 2015) from the sequence of H_C/A5.

Structure Refinement and Analysis

Immediately apparent in the crystal structure was density between the S_γ of Cys_{1280} and N_ζ of Lys_{1236}. Both sidechains were modelled into the density and following a round of refinement there was clearly density for a covalently bound atom between the two side chains. A methyl group was placed in between the S_γ of Cys_{1280} and N_ζ of Lys_{1236}, connecting them. After refinement, the atomic model fit the density extremely well. Some residual negative electron density in the Fourier difference map

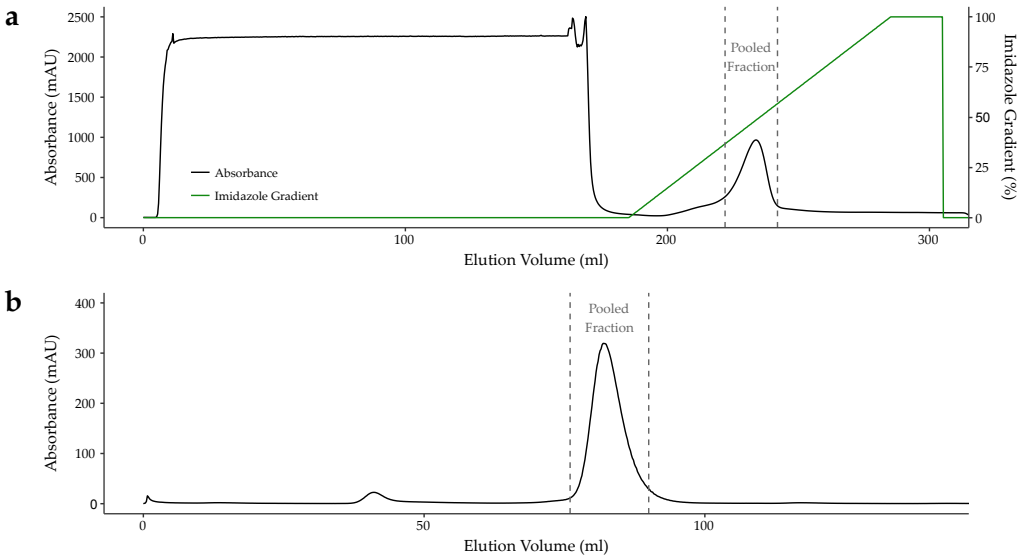


Figure 3.18: Purification of H_C/A5

(a) Chromatogram of the His-tag purification. **(b)** Chromatogram of the gel filtration purification.

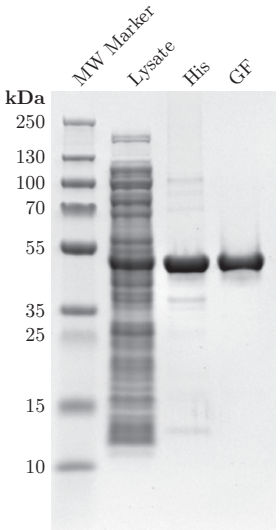


Figure 3.19: H_C/A5 Purification SDS-PAGE.

4 – 12% gradient SDS-PAGE following purification of Hc/A5. Lane 1: Molecular weight marker. Lane 2: Crude lysate before purification. Lane 3: Pooled fractions after His-tag purification. Lane 4: Pooled fractions after gel filtration.

Table 3.8: H_C/A5 Crystallographic Data Collection and Refinement Statistics

	H_C/A5
Data collection	
Beamline	I03 DLS
Wavelength (Å)	0.9763
Crystallographic statistics	
Space Group	$P2_12_12_1$
Cell lengths a,b,c (Å)	43.55, 60.27, 185.15
Cell angles α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	57.36–1.14 (1.16–1.14)
R_{merge}	0.075 (1.093)
R_{meas}	0.084 (1.179)
R_{pim}	0.036 (0.623)
$CC_{1/2}$	0.997 (0.620)
Mean $< I/\sigma I >$	10.7 (1.5)
Completeness (%)	99.8 (97.8)
Total reflections	1,667,857 (61,786)
Total unique reflections	177,879 (8,513)
Multiplicity	9.4 (7.3)
Refinement statistics	
R_{work}/R_{free}	0.139/0.159
RMSD bond lengths (Å)	0.01
RMSD bond angles (°)	1.55
Ramachandran statistics (%)	
Favoured	97
Allowed	3
Outliers	0
Wilson B-factors (Å^2)	11.56
Average B-factors (Å^2)	
Protein	17.10
Water	31.13
No. of atoms	
Protein	3691
Water	456
PDB Code	–

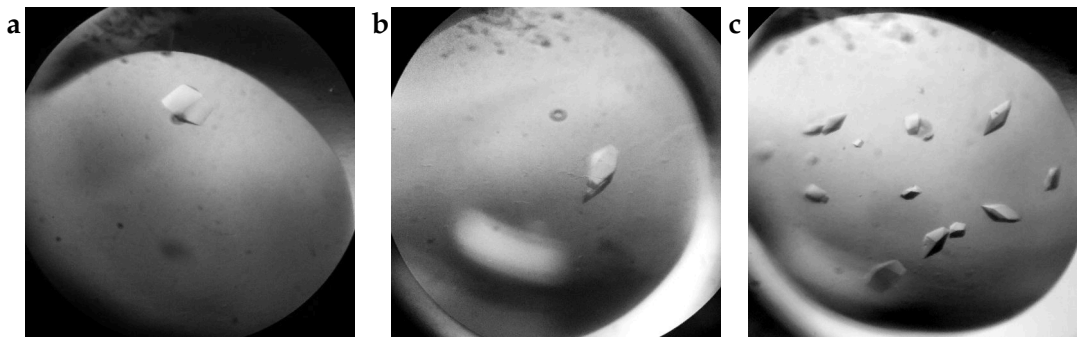


Figure 3.20: H_C/A5 Crystal Morphologies
A single crystal of H_C/A5 initially appeared in **(a)** G5 of The MORPHEUS screen (Molecular Dimensions). Manual optimisation produced further crystals **(b,c)** which failed to produce good quality diffraction data.

(Fo-Fc) was observed around the S_γ atom, while not unusual, this is an indication of some X-ray induced radiation damage. This is not a normal motif found in botulinum neurotoxins and so we looked for further evidence to justify the modelling of such a bond. Fortunately, some weak anomalous data were observed during initial data collection which could be used to generate an anomalous difference map. The anomalous map contained a high level of noise but by decreasing the contour level, the largest of the peaks could be identified which each appeared in the location of sulfur atoms. The sulfur anomalous peaks confirm the locations of each cysteine around the methylene-bridged cysteine and lysine (Figure 3.22)

This specific bond between a lysine and cysteine side chain is quite perplexing although there are some examples of a similar bond identified in previous protein crystal structures – although in most cases it has not been modelled (Ruszkowski and Dauter, 2016). The mechanism surrounding the formation of a methylene-bridge lysine and cysteine is not fully known but Ruszkowski and Dauter (2016) suggest two possible routes. The first would involve the reaction of Lys N_ζ with CO_2 forming a carbamic acid which is susceptible to nucleophilic attack from the Cys S_γ. Then, further dehydration would result in an S-alkyl thiocarbamate bridge which may be reduced to form a methyl-bridge. Another possible mechanism involves a reaction with Lys and the biologically relevant formaldehyde molecule. Lys can be dehydrated by HOCO in an acidic environment to produce an iminium cation which is itself prone to react with nucleophiles such as Cys residues and would result in a methylene

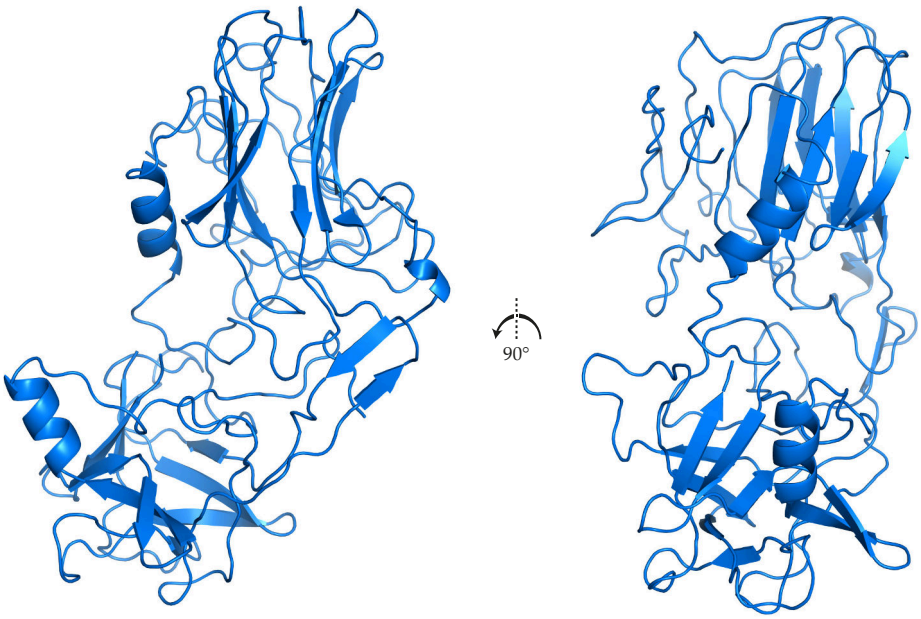


Figure 3.21: H_C/A5 Crystal Structure

Overall cartoon representation of H_C/A5.

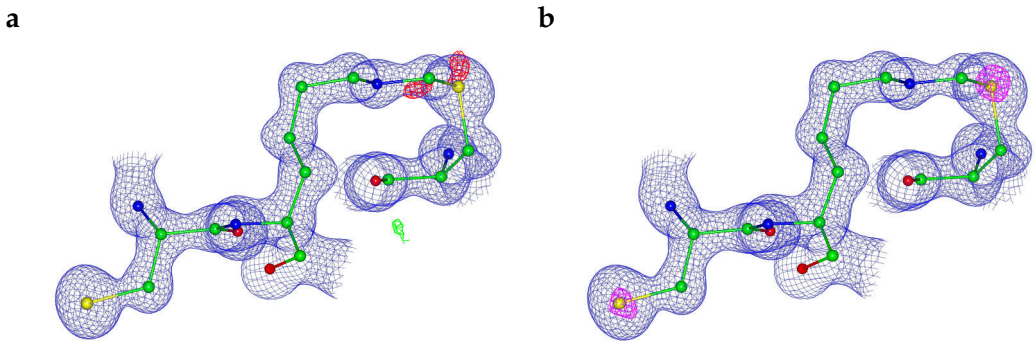


Figure 3.22: A methylene-bridged cysteine and lysine bond in H_C/A5

(a) The 2mFo-DFc map is shown (blue, contoured to 1σ), a small amount of negative electron density is observed in the Fo-Fc difference map (red, contoured to 3σ) map around cys. **(b)** A Composite Omit map (blue, contoured to 1σ) and an anomalous difference map (pink, contoured to 4σ) confirms the location of each sulphur atom.

bridge. Whether this bond is biologically relevant remains to be seen. It is worth noting that other H_C structures have indications of the presence of this bond such as Lys – Cys probe clashes and unmodelled electron density.

3.3.6 H_C/A6 Crystal Structure

Protein Purification

H_C/A6 shares 90.61% identity with H_C/A1 (Table 3.1). The H_C/A6 construct was produced by sub-cloning the H_C domain DNA from a full-length BoNT/A6 sequence (UniprotKB: C9WWY7) and adding an N-terminal histidine tag. The protein was purified in a two step process (Figure 3.23) and the peak corresponding to H_C/A6 was pooled for crystallisation experiments (Figure 3.24). Some lower molecular weight bands were visible directly under the band corresponding to H_C/A6 on an SDS-PAGE gel indicating possibly some degradation or truncated forms (Figure 3.24). These could not be removed by size-exclusion chromatography and crystallisation screening was attempted with this purity.

Crystallisation and Data Processing

Initial BoNT/A6 crystals were grown in condition G7 from Pact Premier (Molecular Dimensions) after two weeks and another condition (A4) produced crystals two weeks later. The crystals from the condition G7 were optimised though manual screening around the hit condition using hanging-drop vapour diffusion. Manual screening resulted in much larger, rod-shaped crystals (Figure 3.25), which were screened for diffraction on beamline I04 at Diamond Light Source. The commercial and optimised conditions all produced good quality crystals with high resolution diffraction (Table 3.9). The best diffraction resulted from an optimised crystal which produced good data to resolution of 1.16 Å (Table 3.10).

Table 3.9: X-ray Datasets Collected for H_C/A6

Screen (Well)	Composition	Crystals (Datasets)	Auto Resolution
PP† (A4)		1 (1)	1.71 Å
PP† (G7)		1 (1)	1.35 Å
Optimised		2 (2)	1.16 Å

† Screen names are abbreviated. PP: Pact premier;

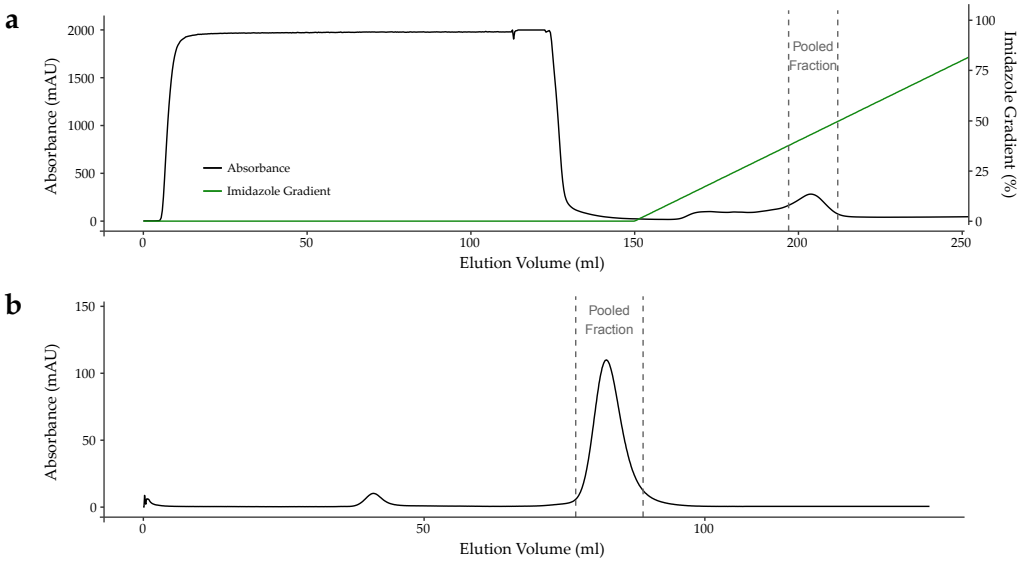


Figure 3.23: Purification of H_C/A6

(a) Chromatogram of the His-tag purification. **(b)** Chromatogram of the gel filtration purification.

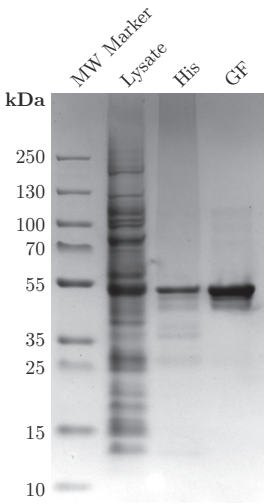


Figure 3.24: H_C/A6 Purification SDS-PAGE.

4 – 12% gradient SDS-PAGE following purification of Hc/A6. Lane 1: Molecular weight marker. Lane 2: Crude lysate before purification. Lane 3: Pooled fractions after His-tag purification. Lane 4: Pooled fractions after gel filtration.

Table 3.10: H_C/A6 Crystallographic Data Collection and Refinement Statistics

	H_C/A6
Data collection	
Beamline	I04 DLS
Wavelength (Å)	0.97634
Crystallographic statistics	
Space Group	$P2_12_12_1$
Cell lengths a,b,c (Å)	39.63, 105.74, 112.86
Cell angles α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å)	77.16–1.16 (1.18–1.16)
R_{merge}	0.101 (1.069)
R_{meas}	0.104 (1.105)
R_{pim}	0.024 (0.276)
$CC_{1/2}$	0.999 (0.626)
Mean $< I/\sigma I >$	18.8 (3.5)
Completeness (%)	98.6 (95.7)
Total reflections	5,676,108 (233,693)
Total unique reflections	162,192 (7,703)
Multiplicity	35.0 (30.3)
Refinement statistics	
R_{work}/R_{free}	0.162/0.201
RMSD bond lengths (Å)	0.02
RMSD bond angles (°)	2.05
Ramachandran statistics (%)	
Favoured	96.89
Allowed	2.87
Outliers	0.24
Wilson B-factors (Å^2)	10.45
Average B-factors (Å^2)	
Protein	17.82
Water	27.49
No. of atoms	
Protein	3560
Water	407
PDB Code	–

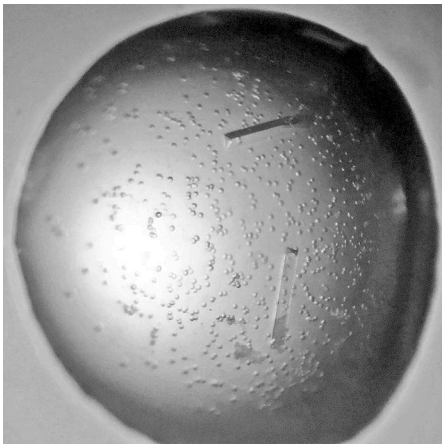


Figure 3.25: H_C/A6 Crystal Morphology
Crystals of H_C/A6 grew into long individual rods in crystallisation drops which also produced many small spherulites.

Structure Refinement and Analysis

The structure was solved by molecular replacement using a model produced on the Phyre2 web server from the H_C/A6 sequence (Kelley *et al.*, 2015). A single protein molecule was present in the asymmetric unit and electron initial electron density was very clear. Surprisingly, the crystal structure of H_C/A6 also revealed interesting density around the S_γ of Cys_{1280} and N_ζ of Lys_{1236}, similar to what was observed for H_C/A5. In this structure however, there appeared to be no methylene-bridge, instead the Cys S_γ and Lys N_ζ appear covalently bound to each other (Figure 3.28a).

There is a large negative peak present in the Fo-Fc difference map for Cys_{1235} indicating some level of radiation damage (Figure 3.28a). To discover whether X-ray induced radiation damage is causing this bond to form, the diffraction data were reprocessed to only include the first 10% of images (100°). As the data were collected using a low transmission during high multiplicity collection strategy, cutting the extra images increases the noise of the data (outer shell $CC_{1/2}$ = 0.226) and the completeness drops to 87% overall. The resulting electron density reveals weaker density for the Cys–Lys bond but it is still clearly present (Figure 3.28b).

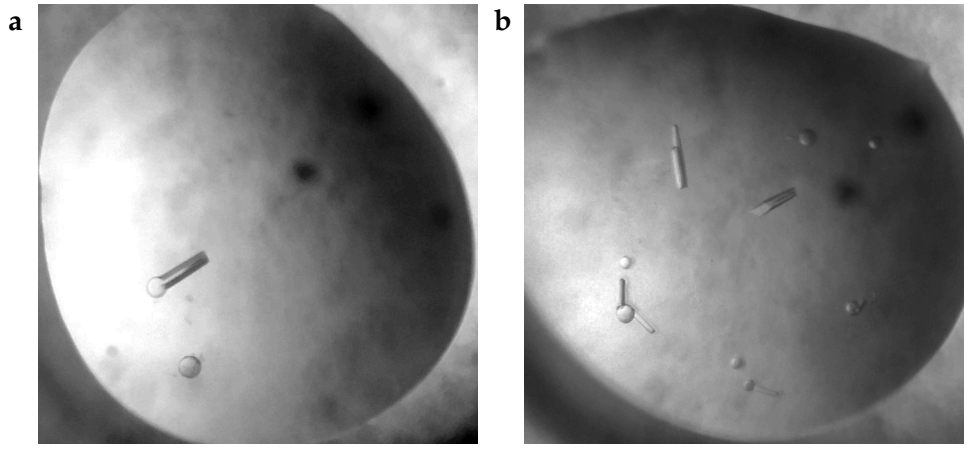


Figure 3.26: H_C/A6–GD1a Crystal Morphology

Crystals which grew following screening of H_C/A6 in combination with equimolar GD1a produced small rod-like crystals which grew out of spherulites.

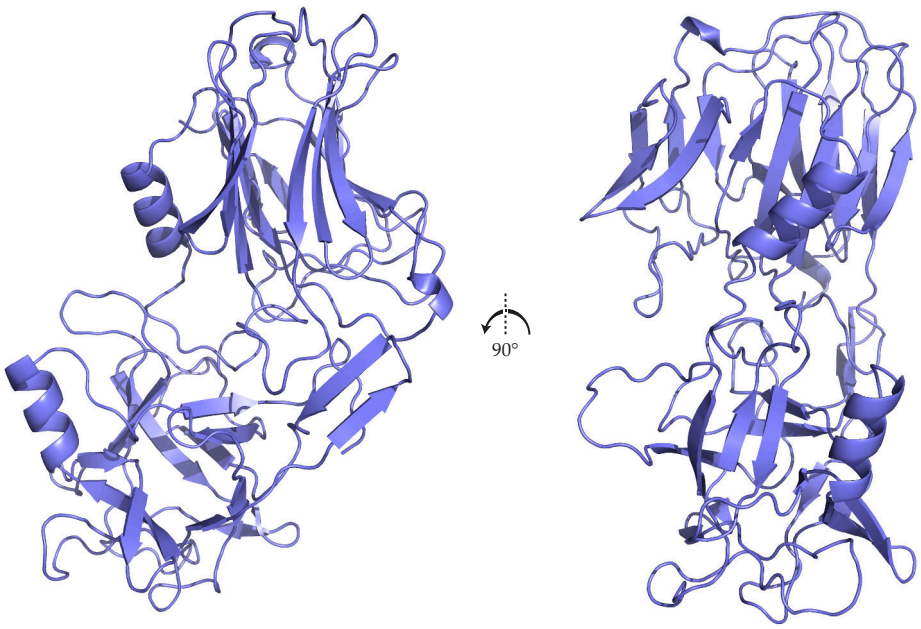


Figure 3.27: H_C/A6 Crystal Structure

Overall cartoon representation of H_C/A6.

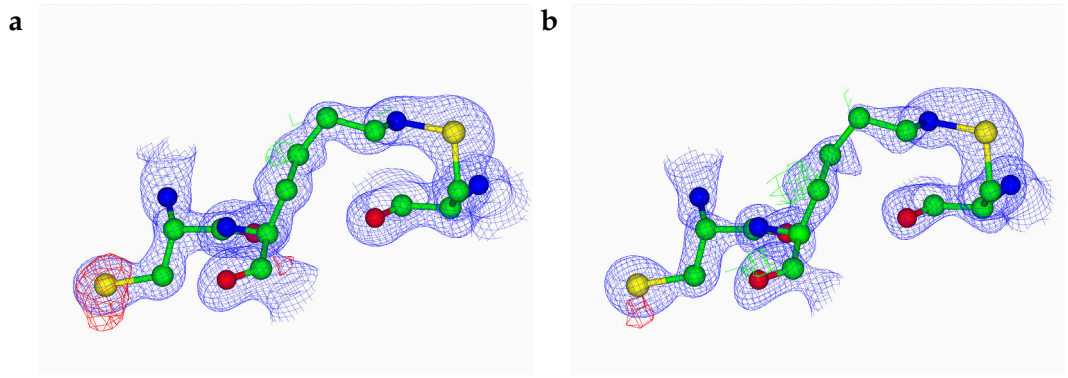


Figure 3.28: A cysteine – lysine bond in H_C/A6

(a) The 2mFo-DFc map is shown (blue, contoured to 1σ), a small amount of negative electron density is observed in the Fo-Fc difference map (red, contoured to 3σ) map around Cys_{1235}. **(b)** The electron density map was recalculated using only the first 1000 images (100° of data) revealing less negative density.

3.3.7 H_C/FA Crystal Structure

Protein Purification

BoNT/FA is a hybrid toxin whose H_C domain shares 84% identity with H_C/A1 and a LC and H_N domain similar to BoNT/F5. The H_C/FA construct was produced by first identifying the H_C domain through alignment with H_C/A1 and then sub-cloning the H_C domain DNA from a full-length BoNT/FA sequence (NCBI: WP_047402807.1), and adding an N-terminal histidine tag. The protein was purified in a two step process in the same manner as the H_C/A subtype domains (Figure 3.29) and the peak corresponding to H_C/FA was pooled for crystallisation experiments.

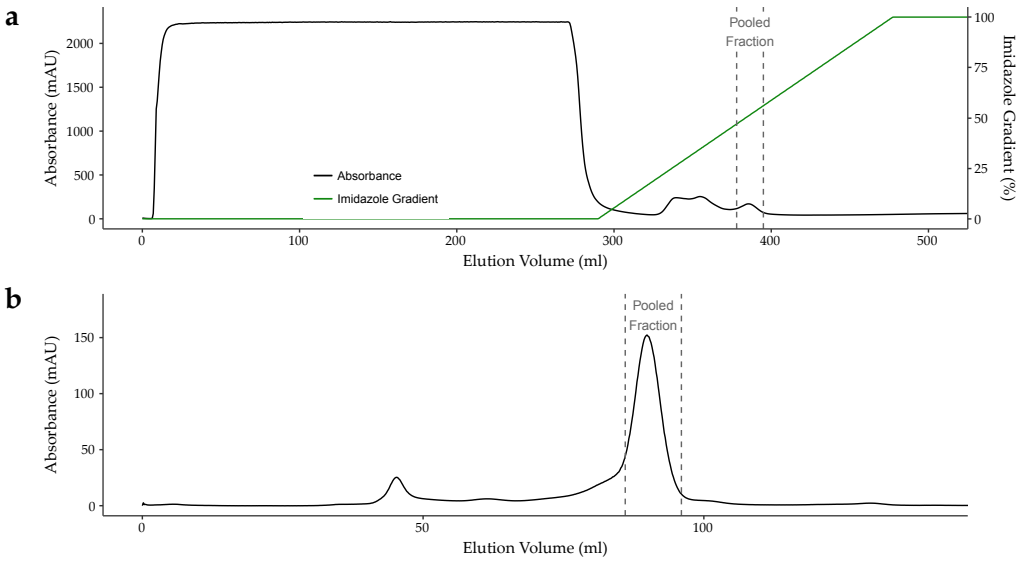


Figure 3.29: Purification of H_C/FA

(a) IMAC purification of H_C/FA resulted in a multiple small peaks of eluted protein. Clarified *E. coli* lysate was loaded onto a HisTrap column and the column washed before a gradient elution. **(b)** Size exclusion chromatography removed the residual contaminants and imidazole from the IMAC purificaiton.

A crystallisation condition was identified which yielded crystals of H_C/FA in the space group P422. A high-multiplicity dataset was collected containing 360 degrees of data over 3,600 images. No significant radiation damage was observed over the course of data collection and thus all data were used. The $CC_{1/2}$ value for the outer shell was

0.832 using a resolution cut-off of 1.95 Å, indicating there was still usable data at this resolution (Evans and Murshudov, 2013) (Table 3.10). The structure was solved by molecular replacement, using a model produce on the Phyre2 web server, and phasing located two molecules related by non-crystallographic symmetry in the asymmetric unit. During preparation of our H_C/FA manuscript, Yao *et al.* (2017) published the same crystal structure (PDB ID: 5V38). Both molecules superpose well with a root mean squared deviation (RMSD) value of 0.35 Å between all atoms.

The crystal structure of H_C/FA is highly similar to the existing structure (PDB ID: 5V38, Yao *et al.*, 2017) and each molecule in the asymmetric unit has an RMSD with this structure from 0.54 Å to 0.36 Å (Figure 3.30). Unlike the existing structure, our data contained density for a previously unresolved loop (R_{1261} – R_{1268}) which is likely involved in ganglioside-binding (Figure 3.31). Yao *et al.* (2017) suggested that the lack of density observed for R_{1261} – R_{1268} loop was due to high flexibility, which is consistent with the high B-factors (74 $Å^2$) we observed in this region relative to the rest of the protein.

Table 3.11: H_C/FA Crystallographic Data Collection and Refinement Statistics

	H_C/FA
Data collection	
Beamline	I03 DLS
Wavelength (Å)	0.97625
Crystallographic statistics	
Space Group	P422
Cell lengths a,b,c (Å)	90.00, 90.00, 90.00
Cell angles α, β, γ (°)	118.00, 118.00, 118.00
Resolution (Å)	24.40–1.95 (1.98–1.95)
R_{merge}	0.201 (1.686)
R_{meas}	0.208 (1.760)
R_{pim}	0.056 (0.502)
$CC_{1/2}$	0.999 (0.832)
Mean $< I/\sigma I >$	12.6 (2.3)
Completeness (%)	100.0 (100.0)
Total reflections	
Total unique reflections	
Multiplicity	26.0 (23.7)
Refinement statistics	
R_{work}/R_{free}	0.180/0.209
RMSD bond lengths (Å)	0.007
RMSD bond angles (°)	0.89
Ramachandran statistics (%)	
Favoured	96.7
Allowed	3.4
Outliers	0
Wilson B-factors (Å^2)	24.55
Average B-factors (Å^2)	
Protein	28.2, 30.0
Water	33.8
No. of atoms	
Protein	6977
Water	609
PDB Code	5MK8

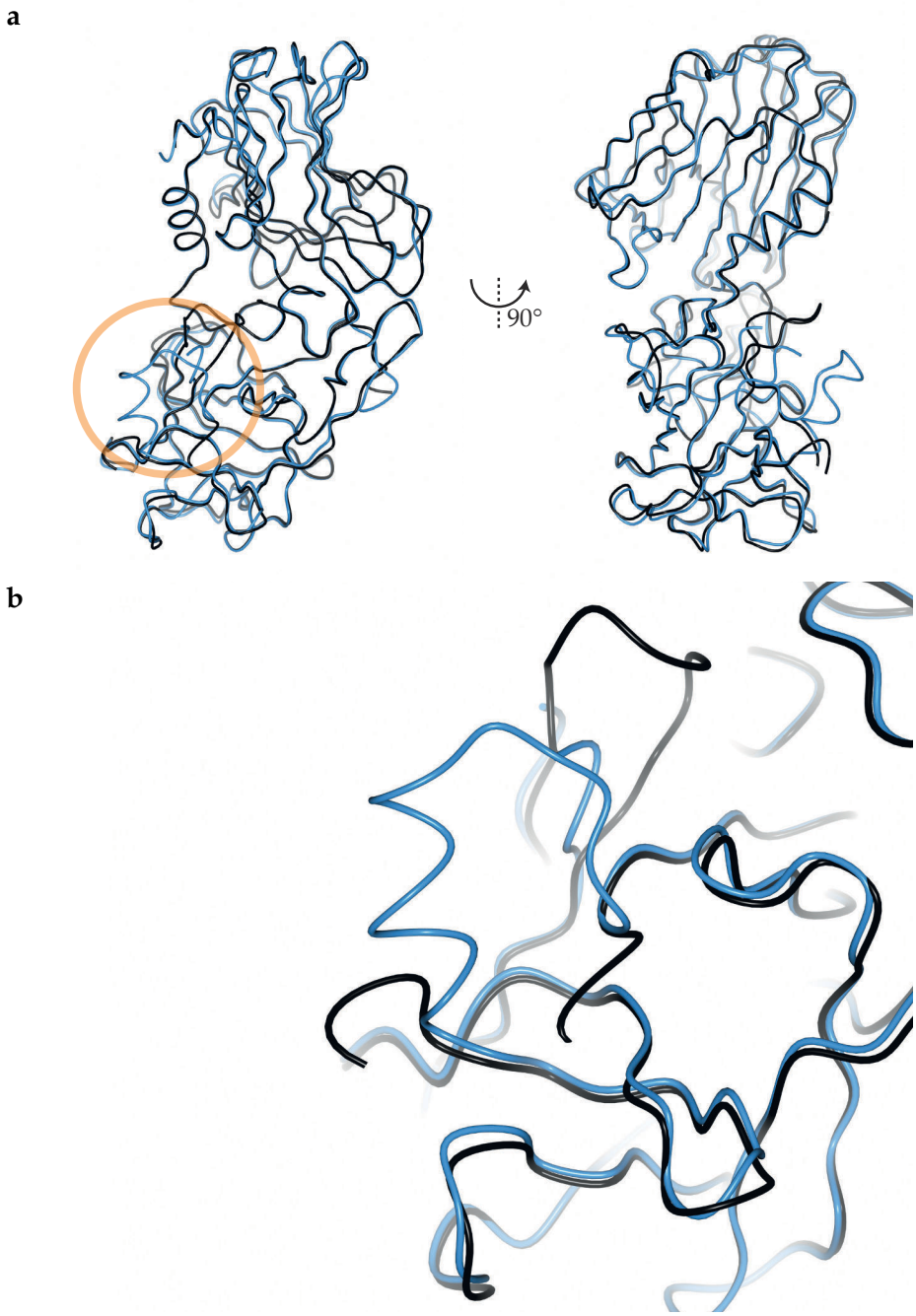


Figure 3.30: Comparison Between H_C/FA Crystal Structures

(a) Our crystal structure of H_C/FA (blue, PDB ID: 5MK8) is superposed onto an existing structure of H_C/FA (black, PDB ID: 5V38). Both structures are displayed as a wire representation of the backbone. Circled in orange is the inferred ganglioside binding site and a zoomed image of this region **(b)** is provided for clarity.

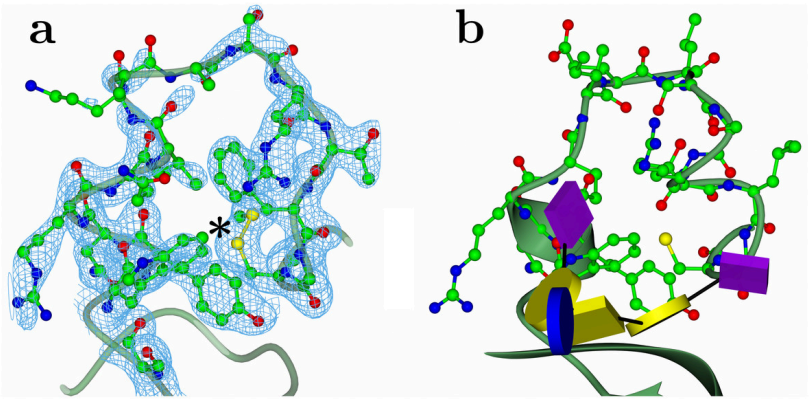


Figure 3.31: H_C/FA Ganglioside Binding Site

(A) Electron density from a composite omit map for H_C/FA. The location of a disulphide bond is marked with an asterisk. **(B)** The equivalent loop from H_C/A1 (2VU9) with GT1b shown in glycoblock representation. Map produced using Phenix package (Terwilliger *et al.*, 2008). Figure reproduced from Davies *et al.* (2018a).

3.3.8 Comparison of H_C/A Crystal Structures

Ganglioside Binding Site

The ganglioside binding site of BoNT/A subtypes is particlarly flexible and in many structures (BoNT/A2, BoNT/A3, BoNT/A6) the electron density for the 1292–1280 loop cannot be fully resolved (Figure 3.32). As identified in the H_C/A1–GT1b crystal structure, the ganglioside interacts with many residues, some of which hydrogen bond while others act to form a hydrophobic pocket. A particularly important and well conserved residue is W_{1266} which forms an electron stacking interaction with the Gal4 moiety of the ganglioside. There is one key variation which appears across the subtypes, that is $Y_{1117}F$ which appears in BoNT/A2, /A3, and /A6. Data from our H_C/A3–GD1a crystal structure indicates that this variation is tolerated and does not seem to have an appreciable effect on ganglioside binding (Figure 3.13). It may however, cause a decrease in binding affinity as multiple hydrogen bonds are lost with the Sia5 moiety of the ganglioside. Our H_C/A crystal structures provide a great starting point for the engineering of H_C domains with greater affinity for ganglioside binding.

SV2 Binding Site

The SV2 binding site of H_C/A subtypes can be separated into two parts, firstly the binding interactions with the SV2 backbone (Figure 3.33), and secondly the binding interactions with the SV2 glycan (Figure 3.34). Many of the H_C/A subtypes contain sequence variation along the β-strand which forms backbone interactions with the SV2 molecule. Variations such as $M_{1144}V$ in H_C/A2, and $S_{1142}N$ in H_C/A4 may be tolerated as the side chains do not form direct interactions. In the case of H_C/A2, a crystal structure of the protein bound to SV2 does indeed show this mutation is tolerated (Benoit *et al.*, 2017). Other mutations may have a much greater effect on the affinity toward SV2. H_C/A1 is the only subtype to contain an Arg at position 1156 which, in the crystal structure bound to glycosylated SV2 (Yao *et al.*, 2016), appears to form a π-stacking interaction with a Phe of SV2. All other H_C subtypes contain a Met at this position, with the exception of H_C/A2 which contains a Glu. Variations such as these do not appear to effect the local fold but may have a strong effect on

binding affinity. Another interesting residue is R_{1249} of BoNT/A1 which extends over the SV2 protein forming a direct interaction with a second β-strand of SV2. In the case of BoNT/A6 the residue is instead a Lys which may still form a hydrogen bond with SV2. The other BoNT/A subtypes contain the $R_{1294}S$ variation, in the crystal structures the Ser residue does not extend as far as the Arg and would likely not be involved in any SV2 interaction (Figure 3.33) – this is also the case as observed in the H_C/A2–SV2 crystal structure (PDB ID: 5MOY) (Benoit *et al.*, 2017).

The Cys_{1235} and Cys_{1280} Disulfide Bond

From the initial crystal structure of full-length BoNT/A1 (PDB ID: 3BTA) determined by Lacy *et al.* (1998) it was found that two cysteine residues within the H_C are capable of forming a disulfide bond. The role of this disulfide is unknown but one would assume a structural role whereby the covalent bond acts to stabilise the fold. Structures determined of the H_C domain of BoNT/A1 since the first full-length structure have revealed a range of possible conformations of the cysteine residues in question.

The identification of a condition that produces two crystal forms of H_C/A1 has prompted further questions regarding the role of the Disulfide bond. Namely, are two distinct species of protein produced during protein expression or is the redox state of this bond sensitive to its environment over time? Considering their location near the ganglioside-binding pocket, and conservation across BoNT subtypes, the redox status of these conserved cysteines may have implications in BoNT stability and manufacture. Botulinum neurotoxins are used therapeutically for many indications and their production is currently from the native host *Clostridium botulinum*. However, considering the safety implications and the advent of engineered BoNT derivatives, such as Targetted secretion inhibitors (TSIs) (Masuyer *et al.*, 2014), production from a recombinant source would be highly desirable. Of the other crystal structures that we have determined, H_C/A2, and H_C/A4 also contain the equivalent disulfide bond. Understanding the role of this bond is very important and further work is required to investigate any impact it may have on the BoNT function.

In the H_C/A5 crystal structure, instead of a disulfide bond, Cys_{1280} forms a bond with Lys_{1236} through a methylene bridge. This highly unusual bond has been previously

observed in other unrelated crystal structures but its significance and the mechanism of bond formation are little understood (Ruszkowski and Dauter, 2016). Confusing matters further, the crystal structure of H_C/A6 contains a bond between Cys_{1280} and Lys_{1236} but without a methlyene bridge, *i.e.* an N–S bind which is highly unusual and likely unstable. Cysteine is regarded as a reactive amino acid and modifications including oxidation but the Lys-CH_2-Cys or even Lys-Cys modification is not well known or understood. It is plausible that these bonds are artefacts of the reducing power of synchrotron radiation and may not have biological relevance. However, they require further investigation to uncover whether or not these interactions are present in solution.

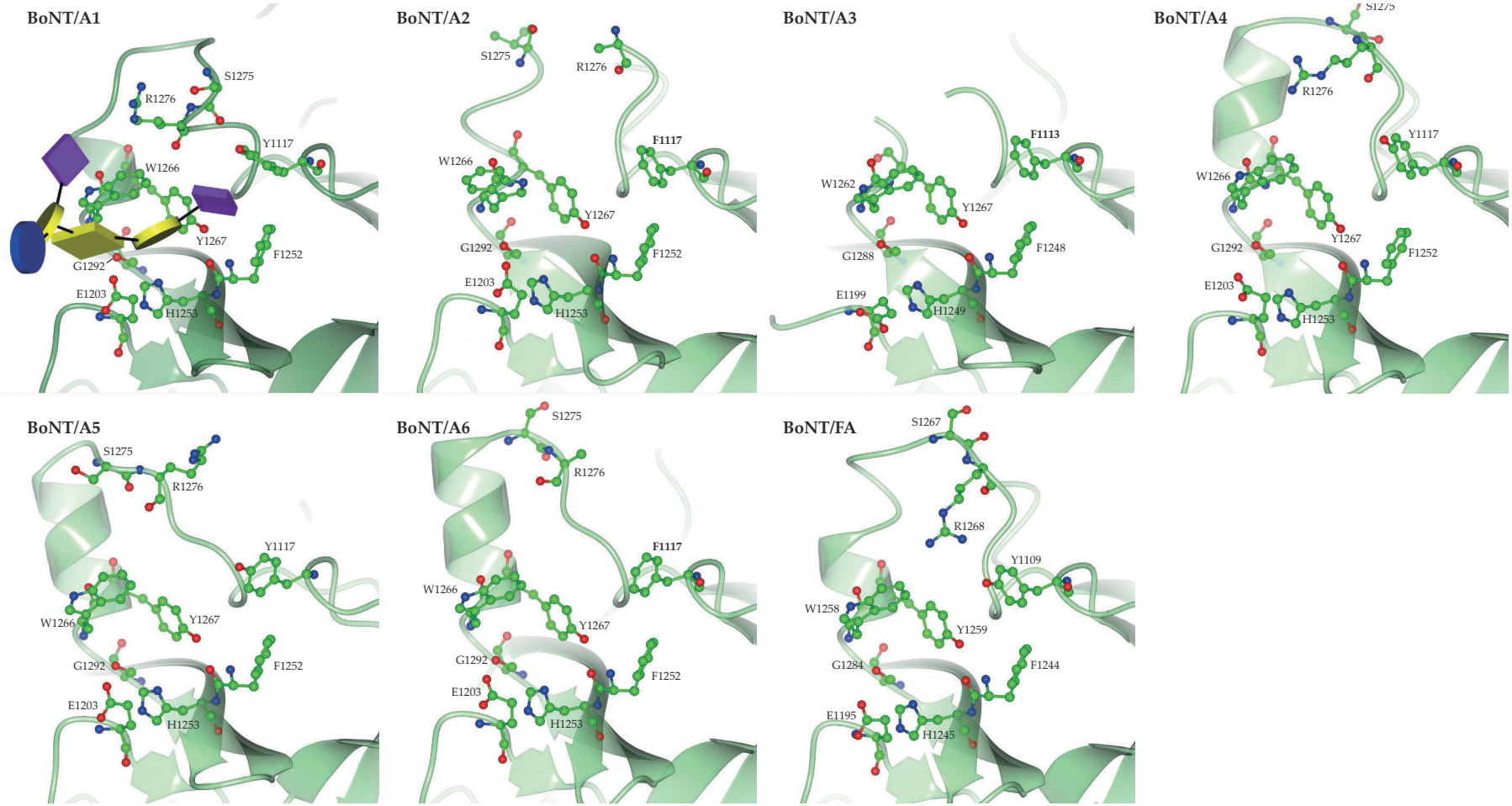


Figure 3.32: BoNT Ganglioside Binding Site

Displayed is the ganglioside binding site of different BoNT/A subtypes. For each H_C the protein is represented in cartoon form with residues that may form interactions with the ganglioside displayed in ball-and-stick format. For BoNT/A1 the crystal structure of the H_C in complex with GT1b is shown. The ganglioside is represented in glycoblock form.

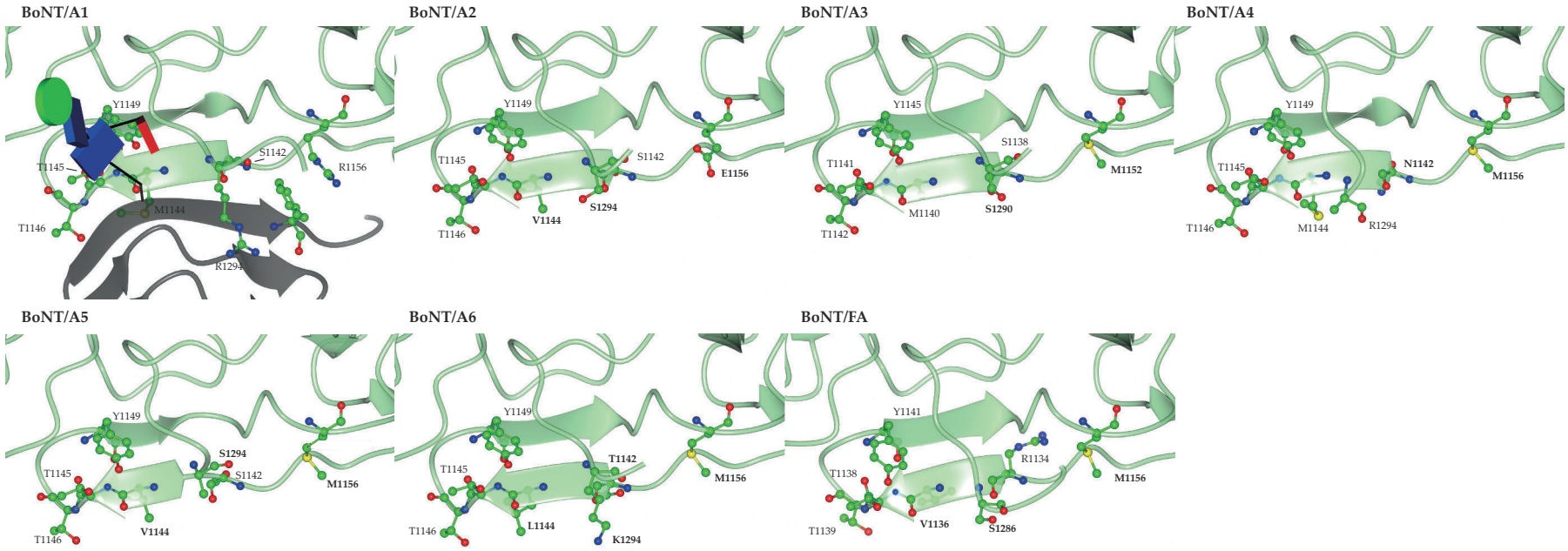


Figure 3.33: BoNT SV2 Binding Site The SV2 binding site of different BoNT/A subtypes. For each H_C the protien is represented in cartoon form with residues that may form interactions with SV2 displayed in ball-and-stick format. For BoNT/A1 the crystal structure of the H_C in complex with glycosylated SV2 is shown. SV2 is coloured black and the glycan represented in glycoblock form.

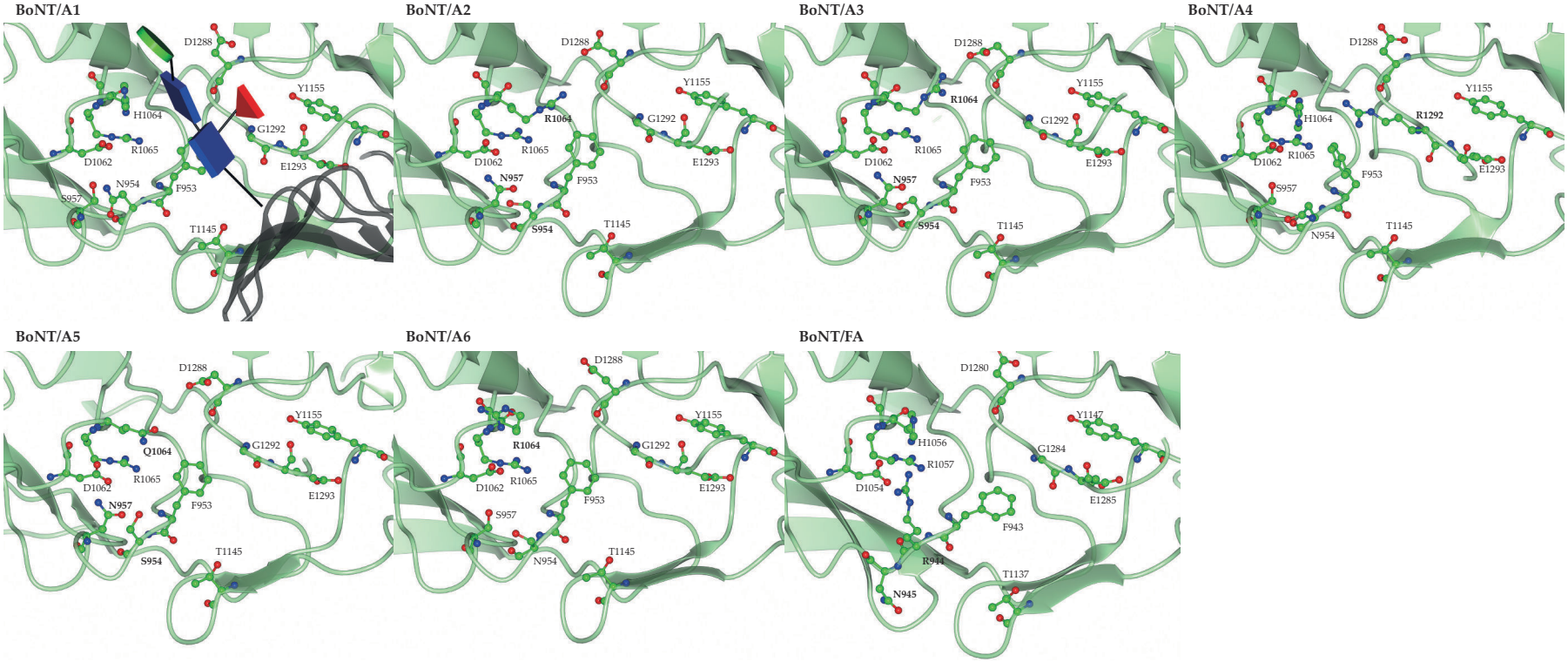


Figure 3.34: BoNT SV2 Glycan Binding Site

Displayed is the SV2 glycan binding site of different BoNT/A subtypes. For each H_C the protein is represented in cartoon form with residues that may form interactions with the SV2 glycan displayed in ball-and-stick format. For BoNT/A1 the crystal structure of the H_C in complex with glycosylated SV2 is shown. SV2 is coloured black and the glycan represented in glycoblock form.

3.4 Conclusions

The work presented in this chapter has resulted in nine crystal structures including the highest resolution BoNT H_C domains determined to date. Previous structural studies have focussed solely on H_C/A1 or H_C/A2 and the data presented here reveals for the first time the precise atomic details of further BoNT/A subtypes. Subtypes from the BoNT/A serotype bind to, or are proposed to bind to, a dual-receptor complex on the cell surface. Investigation of the binding sites through sequence and structural analysis reveals many amino acid variations across the subtypes which may cause differences in binding-affinity to each receptor. Our high resolution models further add to the body of knowledge around BoNT receptor binding and will increase the available information when exploring the possibility of future engineered BoNTs.

CHAPTER 4

Characterisation of SV2 Binding

4.1 Introduction

There are seven distinct serotypes of botulinum neurotoxins (BoNTs) termed BoNT/A through BoNT/G (Coffield *et al.*, 1997). Each serotype can be further categorised into subtypes based on amino acid sequence identity. In the BoNT/A serotype there are eight subtypes (BoNT/A1–/A8) which share between 84% and 97% sequence identity. One of these subtypes, BoNT/A1 is a widely used and commercially available therapeutic used to treat a number of conditions including cervical dystonia, overactive bladder, strabismus, and glabellar lines (Kane *et al.*, 2015; Fonfria *et al.*, 2018; Johnson, 1999). BoNTs utilise an incredible technique for targetting neuronal cell specifically through the dual-receptor complex. BoNT/A binds to a dual-receptor complex consisting of a ganglioside and synaptic vesicle glycoprotein 2 (SV2) on the neuronal cell surface during intoxication (Dong *et al.*, 2006; Yao *et al.*, 2016; Hamark *et al.*, 2017). The crystal structure of H_C/A1 in complex with glycosylated SV2 was determined by Yao *et al.* (2016) and it reveals an array of interactions with the SV2-glycan. Previous data had only shown interaction with the SV2 backbone, an interaction which would not account for the extreme toxicity of the BoNT (Benoit *et al.*, 2014). A detailed understanding of the structural aspects of the BoNT subtypes will prove useful in identifying regions of variability that may help uncover conserved mechanisms of binding. In addition to structural work, attempts to accurately measure the binding affinity to each receptor will in turn help to guide efforts in developing novel BoNT therapeutics. Despite the specificity, it is still desirable to produce a therapeutic BoNT with the greatest possible binding affinity and efficiency, which would avoid the spread of the toxin away from the application site (Kerscher *et al.*, 2012; Tao *et al.*, 2017).

Establishing a robust technique for analysing the binding affinity of BoNTs to their receptors is highly desirable. Not only would it enable better characterisation of existing BoNTs but it could also be used to characterise novel BoNTs which may have engineered H_C domains to improve affinity to target receptors.

4.1.1 Biolayer Interferometry

One technique for analysing binding interactions between a protein and a target-receptor is biolayer interferometry (BLI). BLI is an optical technique based on reflectometric interferometry spectroscopy (RIfS) that analyses the interference pattern of light that has been reflected from two surfaces. In the case of BLI, light is sent down an optical fibre and reflects back to the detector. When protein is bound to the sensor tip, light reflects back from not only the internal surface but also from the layer of protein causing the returned light to have a measurable interference pattern (Concepcion *et al.*, 2009) (Figure 4.1).

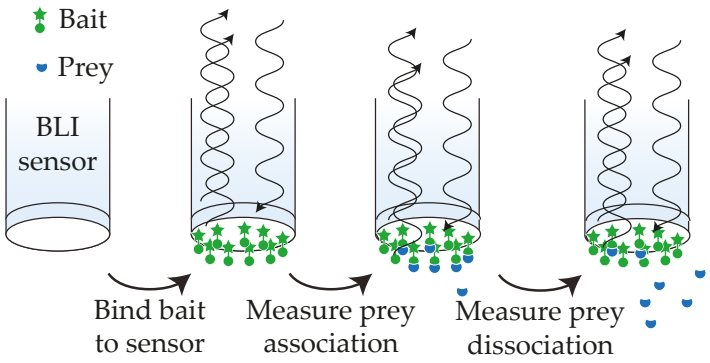


Figure 4.1: Biolayer Interferometry
Measurement of protein binding constants using Biolayer interferometry (BLI) requires the bait protein to bind to the sensor end. Following this the association and dissociation of the prey protein can be measure by recording the light reflected from the sensor tip.

To analyse protein–protein interactions, the BLI sensor must first bind to one of the proteins (bait). The sensor tip can be coated in an antibody-containing matrix which allows targetting of specific proteins, for example an anti-glutathione-S-transferase (anti-GST) sensor tip will selectively bind to GST-tagged proteins. After binding of the first protein, the resulting interference signal can be normalised and the second protein (prey) allowed to bind. The additional layer of protein from the second binding event can be measured and processed to provide association parameters. The sensor can then be transferred into a solution containing no protein and the dissociation

parameters also measured.

Presented here is an investigation into the feasability of utilising BLI for analysing the interaction between the BoNT H_C domain and a protein-receptor, specifically glycosylated SV2C.

4.2 Methods

All materials were purchased from Sigma Aldrich unless otherwise specified.

4.2.1 Construct Cloning and DNA Production

DNA encoding human SV2C luminal domain 4 was provided by Ipsen Bioinnovation Ltd. The final construct for mammalian cell expression was produced in multiple stages. First a GST-tag was added to the N-terminal end of the SV2C sequence through sub-cloning into the pOPINJ vector from the Oxford Protein Production Facility (OPPF) (Berrow *et al.*, 2007). Next, an N-terminal IL2 signal sequence was added before the GST-tag by PCR amplification of GST-SV2C using a primer encoding IL2. Finally, the entire construct was ligated into linearised pOPINJ which lacked any gene insert. The final plasmid was confirmed by sequencing (MWG Eurofins, Germany) and transformed into *E. coli* Top10 cells for production of a large quantity of DNA for transfection. 500 mL cultures were grown and processed by alkaline lysis using a GigaPrep kit (Qiagen) according to manufacturers instruction to isolate the DNA.

4.2.2 HEK293T Transient Transfection

Transfections were performed in two expanded surface (2125 cm^2) roller bottles (Cellmaster™, Greiner Bio-One). Each bottle was seeded using HEK293T cells from a T175 flask at $>$ 90% confluency. Briefly, media was removed from a T175 flask and cells washed once using 10 mL PBS. 5 mL trypsin solution was added and the flask was rocked until all cells had dissociated. 20 mL DMEM (Dulbecco's Modified Eagle Medium, ThermoFisher Scientific, 41965039) containing 10% (v/v) FBS was mixed

with the trypsin solution and the total volume (25 mL) added to one roller bottle. The final volume of each roller bottle was 250 mL after addition of 225 mL DMEM (10% (v/v) FBS). Roller bottles were incubated at 37 °C in a 5% CO_2, humidified atmosphere incubator whilst rotating at 0.5 rpm for the first 24 hours post-seeding and 0.8 rpm thereafter. The media was then removed from each roller bottle and replaced with 200 mL DMEM (2% (v/v) FBS). To produce the transfection solution, 100 mL DMEM (2% (v/v) FBS) was divided equally between two vessels. To the first, 1 mg plasmid DNA was added and to the second 1.75 mL of 1 $mg\,mL^{-1}$ polyethylenimine (PEI, Sigma Aldrich, 408727) pH 7.5. Both solutions were combined and vortexed briefly before 10 min room temperature incubation. A 50 mL volume of the transfection solution was subsequently added to each roller bottle which was then incubated at 37 °C in a 5% CO_2, humidified atmosphere incubator whilst rotating at 0.8 rpm. Four days later the media was harvested and stored at -20 °C until needed. 250 mL DMEM (10% (v/v) FBS) was added to each roller bottle which was left to incubate for a further four days. Finally the media was again harvested from each bottle, stored at -20 °C and the cells discarded.

4.2.3 Protein purification

Raw cell media was purified in batches of 500 mL using the following method. Cell media was centrifuged at 4000 g for 30 min and filtered through a 0.2 µm filter to remove cell debris. An XK 16/20 (GE Healthcare) column was packed with 10 mL Glutathione Sepharose 4B (GE Healthcare). The column was equilibrated using phosphate buffered saline (PBS; Oxoid, BR0014). Cleared cell media was loaded onto the column using a flow rate of $0.5\,mL\,min^{-1}$. After loading, the column was washed using two column volumes of PBS and protein eluted into 1 mL fractions using 50 mM Tris-HCl, 200 mM NaCl, 20 mM reduced glutathione, pH 8.0. Eluted fractions were analysed by SDS-PAGE (see section 3.2.1). Fractions containing GST-SV2C were buffer-exchanged into 50 mM Tris-HCl, 200 mM NaCl, pH 8.0 and concentrated to 0.5 $mg\,mL^{-1}$ using a 10 MWCO spin concentrator. Finally, protein was stored at -20°C until needed.

4.2.4 Glycosylated hSV2C Mass Spectrometry

20 µg of purified hSV2C was run on a 10% SDS-PAGE gel. The band was excised using a sterile blade and sent to the Discovery Proteomics Facility (Oxford University) for mass spectrometry.

4.2.5 Protein – Protein Binding Assay

BoNT H_C domain to gSV2C binding interactions were assayed using BLI on an OctetQK. GST-tagged gSV2C was diluted to 10 µg mL^{-1} in PBS and 200 added to each well of column B in a 96-well plate (black, Greiner). The BoNT H_C domain to be assayed was diluted in a concentration series to: 600 nM, 300 nM, 150 nM, 75 nM and 37.5 nM. 200 of each concentration was added to respective wells of column D in the 96-well plate. 200 PBS was added to all wells in columns A, C and E. Sensors were pre-soaked in 200 PBS, 10 min prior to starting the assay. A baseline measurement was taken with sensors in column A for 100 s. GST-tagged gSV2C was loaded onto the sensors for 1000 s. A second baseline was measured for 100 s. Sensors were dipped in wells containing BoNT H_C domains for 1000 s before dissociation for another 1000 s.

4.2.6 Analysis

A reference sensor was used to subtract a baseline measurement from all other sensors in the Analyse (ForteBio) software package. A 2:1 heterogeneous model was fitted to the data using a global fit across sensors in both the Analyse software (ForteBio) and GraphPad Prism.

4.3 Results and Discussion

4.3.1 Purification of Glycosylated SV2C

The luminal domain 4 of SV2 isoform C (SV2C) expression was performed in mammalian cells to ensure glycosylation. A large quantity of the SV2 construct DNA (10 mg) was isolated from *E. coli* culture and used to transfect two 2 m^2 roller bottles of HEK293T cells. Expressed protein was secreted by the HEK293T cells and recovered from the cell media. The SV2C was purified from the cell media in a two stage process, first by affinity chromatography with a GSTrap column (GE Healthcare) and secondly by size-exclusion chromatography (Figure 4.2). A very small amount of SV2C was produced per batch indicated by the low protein absorbance levels on the purification chromatograms (Figure 4.2). The GST-tagged protein had a calculated mass of 38.1 kDa and appeared as a smear of slightly higher molecular weight by SDS-PAGE (Figure 4.3). A faint, lower molecular weight, band was also present and presumed to be free GST. A sample sent for mass spectrometry analysis observed the masses of complex glycans attached to both N_{534} and N_{559} of the SV2 protein, the latter is involved in H_C/A1 binding interactions (Yao *et al.*, 2017).

4.3.2 Recording of Binding Data

H_C/A1–/A6 were previously purified as presented in Chapter 3. Matching buffers throughout the experiment are essential. All proteins were buffer exchanged into a commercially available formulation of phosphate buffer saline (PBS) pH 7.4 before binding experiments. To ensure stable readings when running assays using the Octet (ForteBio), the equipment was switched on at least two hours before starting a run for the temperature to fully stabilise. Running an assay before temperature stabilisation could cause the baseline to drift significantly. Various concentrations of H_C domains were assayed between 600 nM and 35 nM. At concentrations above 300 nM, non-specific binding was observed and at concentrations below 70 nM the signal-to-noise was very low. For these reasons, three concentrations were used for each assay 300 nM, 150 nM, and 75 nM.

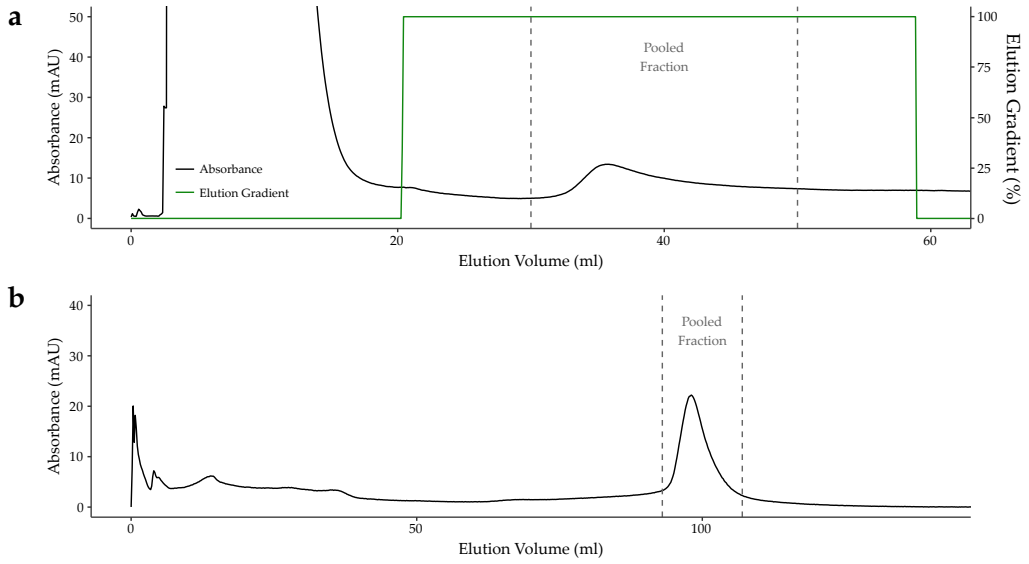


Figure 4.2: Purification of Glycosylated SV2C

(a) Mammalian cell media containing secreted SV2C was purified using affinity chromatography on a GSTrap column (GE Healthcare). **(b)** The protein was pooled and further purified by size-exclusion chromatography.

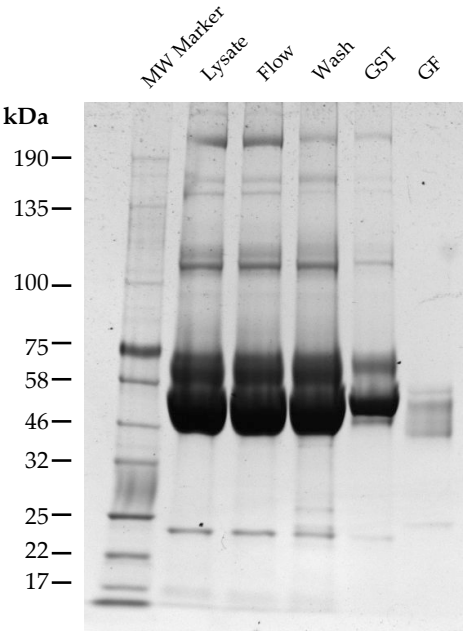


Figure 4.3: Glycosylated SV2C Purification SDS-PAGE.

10% SDS-PAGE following purification of glycosylated SV2C from mammalian cell culture. Lane 1: Molecular weight marker. Lane 2: Crude media before purification. Lane 3: Unbound protein after GSTrap column (GE Healthcare) purification. Lane 4: Column wash. Lane 5: Eluted protein from the GSTrap column. Lane 6: Eluted protein from size-exclusion chromatography.

4.3.3 Processing and Analysis of Binding Data

The post-processing of binding data can have a significant effect on overall values and curve fitting procedures can be quite sensitive to slight differences between datasets. To ensure data between BoNT subtypes was comparable, an identical procedure was followed for each. Data was normalised by reference subtraction. This step can be prone to systematic errors as we are relying on collecting very good data for our reference set. To take an incorrect reference subtraction into account when fitting a model, a parameter was added to provide a global offset to the response values. For example, a small error in background subtraction can be adjusted for with the addition of this parameter. In addition to offset adjustments, each biosensor has its own unique binding capacity due to variances in manufacture. The parameters for R_{max} where not constrained between datasets to account for this. All other parameters were constrained across all datasets. The data clearly did not fit a 1:1 binding model, instead a 2:1 heterogeneous binding model was used. A previous assay which utilised surface plasmon resonance (SPR) to measure the binding of H_C/A1 to SV2 also observed 2:1 binding (Yao *et al.*, 2016). The exact reason for non-1:1 binding is not clear but we speculate that the H_C–glycan interaction and H_C–SV2-protein interaction may occur as two independent events.

Data from three independent runs were collected for H_C/A1 to /A6. Binding models were only able to be fit to collected data from H_C/A1, /A2, /A3, and /A6. We speculate that increased protein aggregation or non-specific binding was causing variation between protein concentrations. Despite repeated runs from multiple batches of H_C/A4 and H_C/A5, stable and consistent binding data could not be achieved. Previous experience of purifying H_C/A4 has shown it to be less stable and more prone to aggregation than other H_C domains (Chapter 3). The 2:1 binding models appear to fit the other datasets well (Figure 4.4) and binding parameters could be calculated (Table 4.1). For each protein, two separate dissociation constants, K_{D1} and K_{D2}, are calculated from each assay. While all K_D values fall within the micromolar to nanomolar range, there is quite some variation across the H_C domains. H_C/A6 in particular, stands out and has a slow dissociation rate compared with the other H_C domains for both binding events in the 2:1 model (Table 4.1).

Yao *et al.* (2016) have previously investigated the binding of $H_C/A1$ to glycosylated SV2 and determined to K_D values of 2.2×10^{-7} M and 1.5×10^{-8} M which are near to the values determined for $H_C/A1$ here. Unfortunately, no other published binding data is available for $H_C/A2$ to $/A6$ and so we cannot fully validate our results.

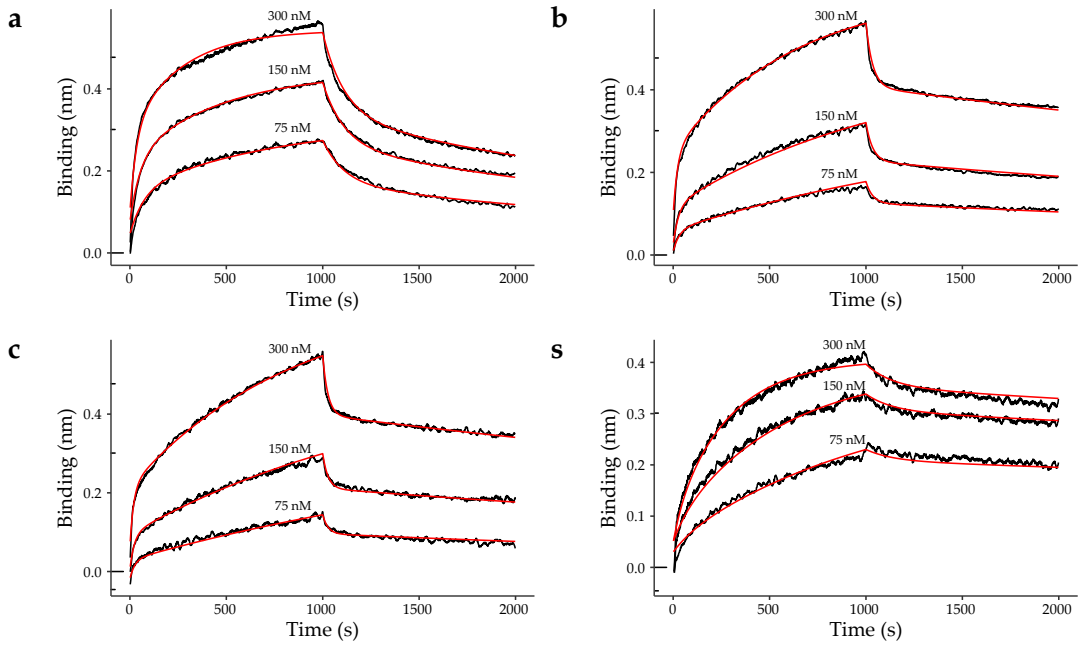


Figure 4.4: H_C/A–SV2 Binding

Binding curves (black) for **(a)** $H_C/A1$, **(b)** $H_C/A2$, **(c)** $H_C/3$, and **(d)** $H_C/A6$ to glycosylated SV2C using Biolayer interferometry. The calculated model is displayed in red. Displayed curves represent the mean from 3 independent experiments. Binding is inferred from the increase in measured optical thickness at the sensor tip in nanometers. The binding curves shown are a combination of association and dissociation events where dissociation begins at Time $= 1000s$.

Table 4.1: H_C/A – SV2C binding parameters

Kinetic binding parameters for the interation between BoNT H_C domains and gSV2C. Displayed values represent the mean from 3 independent experiments and a standard error given for each value.

Subtype	k_{on1} ($M^{-1}s^{-1}$)	k_{off1} (s^{-1})	K_{D1} (M)	k_{on2} ($M^{-1}s^{-1}$)	k_{off2} (s^{-1})	K_{D2} (M)
A1	$1.1 \pm 0.018 \times 10^4$	$5.5 \pm 1.8 \times 10^{-4}$	$5.2 \pm 0.15 \times 10^{-8}$	$6.8 \pm 2.1 \times 10^3$	$9.0 \pm 0.18 \times 10^{-3}$	$1.1 \pm 0.031 \times 10^{-7}$
A2	$5.1 \pm 0.061 \times 10^5$	$2.9 \pm 1.3 \times 10^{-3}$	$5.6 \pm 0.63 \times 10^{-7}$	$4.5 \pm 0.18 \times 10^3$	$2.0 \pm 0.11 \times 10^{-4}$	$4.5 \pm 0.32 \times 10^{-8}$
A3	$4.9 \pm 0.330 \times 10^4$	$3.7 \pm 0.0005 \times 10^{-2}$	$7.4 \pm 0.48 \times 10^{-7}$	$3.3 \pm 0.077 \times 10^3$	$1.9 \pm 0.047 \times 10^{-4}$	$5.7 \pm 0.20 \times 10^{-8}$
A4	–	–	–	–	–	–
A5	–	–	–	–	–	–
A6	$4.6 \pm 0.460 \times 10^4$	$7.5 \pm 69 \times 10^{-7}$	$1.6 \pm 0.14 \times 10^{-7}$	$1.1 \pm 0.0089 \times 10^4$	$7.8 \pm 0.69 \times 10^{-5}$	$6.8 \pm 0.58 \times 10^{-9}$

Binding parameters could not be determined for H_C/A4 or H_C/A5.

4.3.4 2:1 Heterogeneous Binding Model

BLI binding data collected on the Octet (Fortebio) may be processed with the built-in proprietary software. However, reliance on proprietary software limits the ability for scientists to share their methods freely as in most cases the software must be purchased. While the proprietary software in many cases is very good, finding an alternative method to process data should always be investigated. In this Chapter, binding data are presented which conform to a 2:1 heterogeneous binding model. We have written a small software library for analysing heterogeneous binding model which can be run within the open-source and free software, R (R Core Team, 2018). The 2:1 model was written using equations 4.1 to 4.6 and an example can be found in Figure 4.5.

$$k_{obs} = k_{on}[analyte] + k_{off} \tag{4.1}$$

$$R_{eq} = \frac{[analyte]Rmax}{[analyte] + k_D} \tag{4.2}$$

$$K_D = \frac{k_{off}}{k_{on}} \tag{4.3}$$

$$R_{\text{association}} = R_{eq}[1 - \exp(-k_{obs}t)] + R_{eq2}[1 - \exp(-k_{obs2}t)] \tag{4.4}$$

$$R_{\text{dissociation}} = R_0 \exp(-K_{off}t) + R_0 \exp(-K_{off2}t) \tag{4.5}$$

$$R = \begin{cases} R_{\text{association}} & \text{if } t < t0 \\ R_{\text{dissociation}} & \text{if } t \geq t0 \end{cases} \tag{4.6}$$

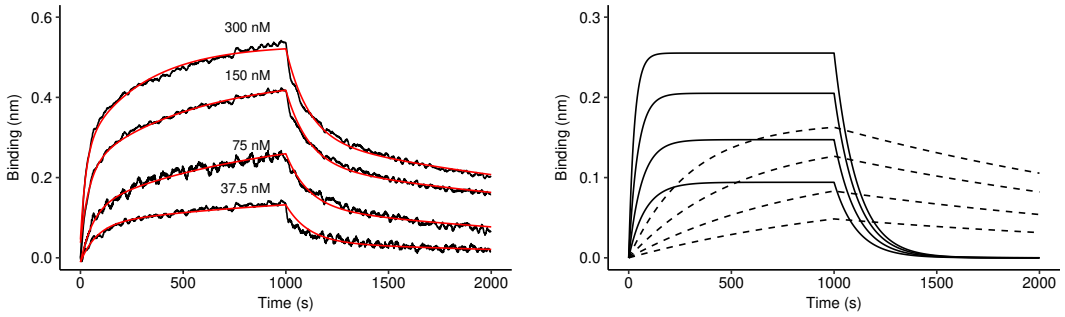


Figure 4.5: 2:1 Heterogeneous Binding Model

Left panels display an example measured 2:1 binding response (black) and fitted 2:1 heterogeneous model (red). Right panels display individual binding components that contribute to the 2:1 heterogeneous binding, solid and dashed lines respectively. Parameters from the fitted model were used to simulate each 1:1 binding response using the R package pbm v1.1.0 (Davies, 2018. Unpublished*).

*Available on The Comprehensive R Archive Network at https://cran.r-project.org/package=pbm

4.4 Conclusions

Accurate measurement of binding parameters is useful for both the development of novel BoNTs and investigation of existing variant. We successfully expressed and purified a glycosylated form of SV2C from mammalian cells and measured binding to a range of BoNT/A H_C domains. The binding data presented here for H_C/A1 are of comparable values to a previously reported binding assay (Yao *et al.*, 2017) suggesting that this is a viable approach to investigate BoNT binding. Two of the proteins assayed, H_C/A4 and H_C/A5, failed to produce data of sufficient quality to fit accurate binding models, possibly due to aggregation or non-specific binding to the sensor tip.

This assay may be useful for a direct comparison of similar variants but there are many other parameters that are worth noting. Most BoNTs form a dual-receptor complex on the neuronal cell surface, and there is evidence of specific serotypes (BoNT/B, /DC and /G) which utilise a loop to add further interaction with the plasma membrane (Fu *et al.*, 2009; Berntsson *et al.*, 2013; Stern *et al.*, 2018). The approach utilised here may not be suitable in all cases as a better understanding of BoNT binding may come from assessment of all binding interactions together. It may be more appropriate to develop an assay which is capable of measuring direct interaction to the cell surface.

CHAPTER 5

Conformation of Botulinum Neurotoxins in Solution

5.1 Introduction

Botulinum neurotoxins (BoNTs) are a family of proteins that cause the condition botulism (Coffield *et al.*, 1997; Schiavo *et al.*, 2000; Rossetto *et al.*, 2014). Traditionally BoNTs have been classified into seven distinct serotypes termed BoNT/A through BoNT/G. In addition several novel BoNTs and BoNT-like proteins have recently been discovered which do not fall under these serotypes. BoNTs are composed of two protein chains, a 50 kDa light chain and a 100 kDa heavy chain, which remain bound by a single disulfide bond. The LC consists of a zinc-dependent endopeptidase while the HC contains a translocation domain (H_N) and a receptor-binding domain (H_C). To date, the full-length crystal structure has been determined for three BoNT serotypes: BoNT/A, BoNT/B and BoNT/E (Lacy *et al.*, 1998; Swaminathan and Eswaramoorthy, 2000; Kumaran *et al.*, 2009). The crystal structure of BoNT/A reveals a unique domain organisation that can be described as a 'butterfly' with both the LC and H_C extending from either side of the central H_N (Lacy *et al.*, 1998). The BoNT/B crystal structure also reveals a similar domain organisation to that of BoNT/A (Swaminathan and Eswaramoorthy, 2000). Unlike both BoNT/A and BoNT/B, the crystal structure of BoNT/E reveals a distinct domain organisation where the H_C wraps around the H_N and LC, resulting in a more compact structural arrangement (Kumaran *et al.*, 2009) (Figure 5.1). BoNT/E has been found to translocate much more rapidly than BoNT/A and BoNT/B (Sun *et al.*, 2012; Wang *et al.*, 2008; Kumaran *et al.*, 2009; Keller *et al.*, 2004).

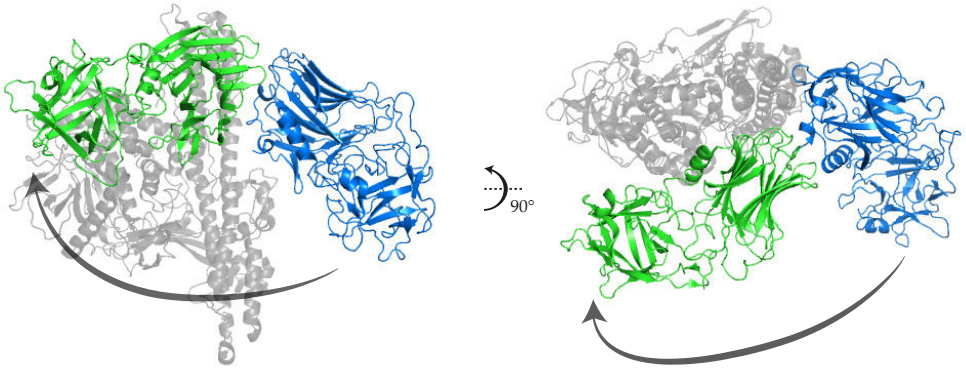


Figure 5.1: BoNT Domain Organisation

A comparison between the BoNT/A H_C (blue) and BoNT/E H_C (green) location with respect to the LC and H_N domains. The LC of both structures were superposed and the LC and H_N of BoNT/A are hidden.

It has been proposed that the increased translocation rate is due to the unique conformation of BoNT/E as determined by X-ray crystallography (Kumaran *et al.*, 2009). The suggestion that conformation of BoNTs can affect translocation speed is particularly interesting for the generation of novel therapeutics. Prior to the release of the BoNT/E crystal structure, BoNT/E was subject to analysis by negative-stain electron microscopy (EM) (Fischer *et al.*, 2008). The EM micrographs revealed the possibility of multiple conformations, one of which compares well with the crystal structure. These data together suggest the protein may be highly flexible and able to adopt multiple conformations in solution. In addition, crystallisation experiments utilise conditions that are designed to favour and stabilise protein-protein interactions resulting in a highly-ordered crystal, and as a result some crystal structures may contain interactions which would not be present under physiological conditions (Dafforn, 2007). As is the case with any X-ray crystal structure, the resulting model is an interpretation of the protein state within a crystal and in some cases further evidence must be sought to confirm any structural analyses.

A low-resolution structural biology technique called small-angle X-ray scattering could be used as a complementary technique to investigate further the full-length structure of BoNTs and determine whether the compact fold of BoNT/E is also its structure in the solution state. This technique, as discussed below, may also be useful for investigating the domain organisation of other BoNTs which do not yet have full-length crystal structures. As discussed in Chapter 2, crystallography enables the high resolution structural determination of any protein, provided that protein can be crystallised in the first place. When investigating different or novel BoNT variants, it is not realistic to expect the protein to crystallise, indeed many attempts of crystallising other BoNT serotypes have so far resulted in little reward.

5.1.1 Small-Angle X-ray Scattering (SAXS)

Small-angle X-ray scattering (SAXS) is a biophysical technique which can be used to characterise biological molecules. In contrast to X-ray crystallography, SAXS experiments do not require crystals and are instead performed on molecules in solution. X-rays interact with individual atoms, generating secondary wavelets which results in coherent scattering. Small-angle scattering results from the coherent scattering of individual atoms within an individual molecule or protein in solution (Jacques and Trewhella, 2010). For biological applications, SAXS is most often used to investigate proteins and other large complexes in solution. The ability to record data from molecules in solution is one of the key benefits of SAXS. When studying proteins, the resulting data can provide information to complement crystal structures including the native oligomeric state, low-resolution solution structure and conformational information (Neylon, 2008). Exposure of the protein solution to X-rays results in X-ray scattering which can be measured using a detector (Figure 5.2).

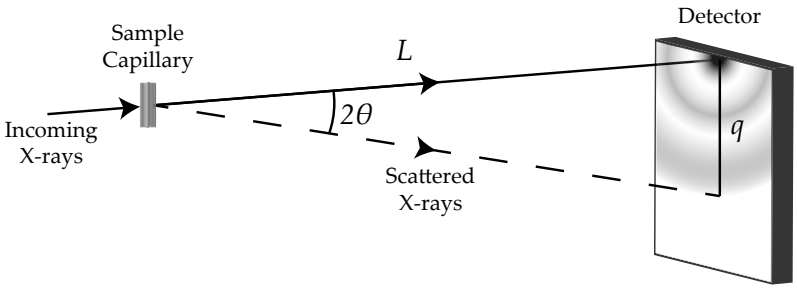


Figure 5.2: SAXS Experiment Schematic The incoming x-ray beam of wavelength, λ passes through the sample in solution within a sample capillary. Some of the x-ray beam is scattered by the sample at the maximum angle 2θ.

In contrast to X-ray crystallography where proteins are arranged in an ordered and well-defined lattice, molecules in solution produce scattered X-rays that are rotationally averaged and at low angles and thus the detector must be positioned far from the sample, usually 5-15 m. The X-ray scattering distance q is measured along with the intensity and these data are combined for subsequent analysis. A typical SAXS

experiment measures the difference in electron density between a protein solution and that of the bulk solvent. Therefore, it is essential that the protein sample is extremely pure and monodisperse to ensure that the resulting data are free of contributions from impurities and aggregates. Protein samples must be reasonably concentrated in order to improve the signal-to-noise of the recorded SAXS signal, and additionally it is helpful if the protein of interest is large as larger proteins will scatter more photons. The photon scattering ability of proteins in solution is extremely weak, for example an estimation using myoglobin (16.7 kDa) concentrated to $1\,\mathrm{mg\,mL^{-1}}$ in water will scatter < 1 in 10^6 photons (Stuhrmann, 1980). Longer exposure times can increase signal but also lead to beam-induced radiation damage of the sample and inaccurate measurements.

The main source of radiation damage is produced after X-ray exposure to the solvent, resulting in free hydroxyl ($OH^{\bullet}$) and hydroperoxyl ($HO_2^{\bullet}$) radicals. These free radicals can attack the peptide chain and drive aggregation. Additionally, having a perfectly matched buffer solution is necessary to subtract the buffer signal from the protein solution. Usually, one would collect data from a protein solution exposure, followed by a buffer exposure. New approaches to improve SAXS data collection have addressed many of the problems mentioned (Trewhella *et al.*, 2017). One such approach is to connect a high-performance liquid chromatography (HPLC) system complete with a size-exclusion chromatography (SEC) column inline with the beamline. This setup, termed SEC-SAXS, can firstly ensure that signal from a highly-pure non-aggregated protein sample is collected in addition to a perfect buffer sample which can be collected before and/or after the protein has eluted. One limitation of using SEC-SAXS is the reduced protein concentration, caused by dilution through the SEC column and associated tubing. Beamlines with low-noise detectors and high flux can overcome this issue but collecting and merging signal from multiple exposures as the eluted protein passed through the beam.

SAXS Data Processing

The first step in processing the data recorded by a typical SAXS experiment is to produce a scatting profile consisting of $I(q)$ versus q, where $q = 4\pi \sin\theta/\lambda$, for both the sample solution and buffer. The buffer scattering is then subtracted and the resulting profile is the scattering of the protein molecule alone. One parameter that can be calculated directly from the SAXS data is the radius of gyration (Rg) (Guinier, 1939). When plotting the natural logarithm of the intensity, $\ln[I(q)]$, against the square of the scattering vector, q^2, the slope of the linear region at low angular range (low q) is directly proportional to the Rg. The Rg can be determined using the Guinier approximation (Equation 5.1).

$$I(q) = I(0)e^{-\frac{1}{3}Rg^2q^2} \tag{5.1}$$

where the intercept at $q = 0$ is the forward scattering at zero angle, $I(0)$. The values of Rg can be determined with high accuracy and changes in Rg can be utilised as a sensitive indicator of the shape of the molecule (Tuukkanen and Svergun, 2014). A non-linear plot at $q.Rg < 1.3$ is indicative of polydispersity, poor background subtraction or unwanted sample interaction effects, either aggregation or repulsion (Blanchet and Svergun, 2013). Protein aggregation results in a large increase of scattering intensity at low q angles which results in the overestimation of both $I(0)$ and Rg. Therefore it is highly important to have monodisperse samples, typically a range of protein concentrations are measured to ensure that a concentration is chosen where aggregation is not an issue. As mentioned above, SEC-SAXS may be able to recover good data where protein samples are likely to aggregate over time. In addition to Rg calculations, an indirect Fourier transform of the one-dimensional scattering profile, $I(q)$ vs. q, can be used to estimate the $Dmax$, that is the largest vector between any two atoms within the protein. Together, these data provide information about the 3-dimensional structure of the protein in solution which can be used to directly compare with existing crystal structures.

5.1.2 Molecular Dynamics

Molecular dynamics is a computational technique used to analyse molecular motions motions including protein folding and ligand interactions (Hug, 2013). The motion of a molecule is simulated *in silico* through solving of Newton's equation of motion simultaneously for all atoms. As a result, larger systems (*i.e.* a large protein) need much more computational power than small systems. To increase the speed and efficiency of such simulations, forces between atoms are calculated using force fields. A force field is a mathematical description of the interaction between specific atomic types in a specific environment that has been empirically determined. This approach can be described as course-grained, where the nucleus and electrons of an atom are treated as one and their motion together follow the rules set out within the force field. Another technique to simulate molecules is by quantum mechanical molecular dynamics, where atoms are explicitly simulated and there is no reliance on a simplified force field. As a result of the increase calculations needed, quantum mechanical simulations are inherently slow compared to classical molecular dynamics and only suited to small molecules or small portions of a larger molecule.

Force fields describe two types of interaction, bonded and non-bonded (Weiner *et al.*, 1984). Bonded interactions generally refers to atoms which are covalently bonded and the force field describes their bond-stretching, bond-rotation and angle-bending propensity. Non-bonded interactions refers to all other interactions such as electrostatic and van der Waals interaction. The flexibility of proteins from simulations can also be quantified following a simulation by calculating the root mean square fluctuation (RMSF) of atoms over the course of the simulation.

5.2 Methods

All materials were purchased from Sigma Aldrich unless otherwise specified.

5.2.1 Construct Production

DNA constructs containing BoNT/A1(0), BoNT/F1(0), and BoNT/FA(0) were provided by Ipsen Bioinnovation Ltd.

5.2.2 *E. coli* Transformation

A single aliquot containing 50 of chemically competent cells was thawed on ice for 10 min. 1 of DNA (up to 200 ng) was added to the cells which were placed back on ice for a further 30 min. Cells were incubated at 42°C for 30 seconds to heat-shock before returning to ice for 5 min. 950 SOC media, prewarmed to 37°C was added to each aliquot which was then placed in an incubator at 37°C while shaking at 225 rpm for 1 hour. 200 from each transformation culture was spread onto a pre-warmed LB (Miller) agar plate containing an appropriate antibiotic. The LB (Miller) agar plate was left at 37°C for 16 hours for colonies to grow.

5.2.3 Protein Expression

A starter culture containing 100 mL mTB was inoculated with *E. coli* from a glycerol stock and incubated at 37°C for 16 hours while shaking at 225 rpm. 10 mL of the starter culture was used to inoculate flasks each containing 500 mL mTB. The flasks were incubated at 37°C while shaking at 225 rpm, when the OD_{600} reached 0.6 the temperature was decreased to 16°C. Protein expression was induced in both flasks when the OD_{600} reached 1 by the addition of IPTG (Isopropyl β-D-1-thiogalactopyranoside) to a final concentration of 1 mM. Following induction the culture flasks were left for a further 16 hours before harvesting cells. Cells were harvested by centrifugation at 4000 xg and cell pastes stored at -80°C.

5.2.4 Purification of Full-Length BoNT/A1(0)

Cell pastes were re-suspended in 10 mL g^{-1} lysis buffer (50 mM Tris-HCl pH 8.0, 25 mM NaCl) plus 2 benzonase (500 units) with stirring for 30 min at 4°C. Re-suspended cells were lysed by passing through a cell disruptor (Constant Systems), using a pressure of 20 kPSI, twice at 4°C. Cell debris was removed by centrifugation at 30,000 x g for 45 min. Clarified lysate was adjusted to contain 1 M ammonium sulfate by the slow addition of conditioning buffer (50 mM Tris-HCl, 3 M ammonium sulfate, pH 8.0). The lysate was incubated at 4°C for 30 min before centrifugation at 30,000 x g for 45 min to remove insoluble precipitate. The clarified lysate was loaded onto a 100 mL butyl sepharose HP column (GE Healthcare). Following sample loading, the column was washed using butyl buffer B (50 mM Tris-HCl, 1 M ammonium sulfate, pH 8.0). Bound protein was eluted using a gradient to 100% butyl buffer A (50 mM Tris-HCl pH 8.0) over 10 column volumes (1 L). Fractions containing BoNT/A1(0), determined by visual inspection of an SDS-PAGE gel (see section 3.2.1), were pooled and concentrated using a 10 MWCO spin concentrator. The pooled and concentrated fractions were desalted into butyl buffer A using a HiPrep Desalting Column (GE Healthcare). Desalted protein was loaded onto a 50 ml Q HP column (GE Healthcare) using a flow rate of 10 mL min^{-1}. The column was washed with butyl buffer A and bound protein eluted using a gradient to 20% Q Buffer (50 mM Tris-HCl, 1 M NaCl, pH 8.0) over 2 column volumes (100 ml).

5.2.5 Purification of Full-Length BoNT/B1(0)

BoNT/B1(0) was purified as described previously (Elliott *et al.*, 2017) and generously provided by Ipsen Bioinnovation Ltd. Briefly, recombinant BoNT/B(0) was expressed using *E. coli* and cells were lysed by ultrasonication in 0.2 M NaCl, 50 mM Tris-HCl, pH 8. Lysate was clarified by centrifugation at 4500 rpm for 30 minutes at 4°C. Clarified lysate was passed through an anion exchange column before buffer exchanging to 1 M $(NH_4)_2SO_4$ in 50 mM Tris-HCl, pH 8. The protein was then purified by hydrophobic interaction resin (high performance butyl Sepharose, GE Healthcare) using a linear gradient to 0.5 M $(NH_4)_2SO_4$ followed by another anion-exchange step. The partially purified protein was then proteolytically cleaved with 0.2 µg mL^{-1} Lys-C to yield

a di-chain and the buffer was adjusted to 1 M $(NH_4)_2SO_4$ in 50 mM Tris-HCl, pH 8. Finally the protein was further purified with a second hydrophobic interaction resin (high performance phenyl Sepharose, GE Healthcare) using a gradient to 0.5M $(NH_4)_2SO_4$ and desalted into PBS pH 7.2.

5.2.6 Purification of Full-Length BoNT/FA(0)

Full-length BoNT/FA(0) was expressed in 2 L of mTB medium. Cell pastes were re-suspended in 10 mL g^{-1} lysis buffer (50 mM Bis-Tris methane-HCl pH 6.8, 0.5 M NaCl) plus 2 benzonase (500 units) with stirring for 30 min at 4°C. Re-suspended cells were lysed by passing through a cell disruptor (Constant Systems) at 4°C. Cell debris was removed by centrifugation at 30,000 x g for 45 min. The lysate was loaded onto a 5 mL HisTrap HP column (GE Healthcare) and eluted using a gradient from 0 M to 0.5 M imidazole-HCl. Finally the protein was further purified with by size-exclusion chromatography (GE Healthcare) using the lysis buffer.

5.2.7 Purification of Full-Length BoNT/F1(0)

Full-length BoNT/F1(0) was expressed in 4 L of mTB medium. Cell pastes were re-suspended in 6 mL g^{-1} lysis buffer (50 mM Tris-HCl, 200 mM NaCl, pH 8.0) plus 2 benzonase (500 units) with stirring for 30 min at 4°C. Re-suspended cells were lysed by passing through a cell disruptor (Constant Systems) twice at 4°C. Cell debris was removed by centrifugation at 30,000 x g for 45 min. Clarified lysate was adjusted to contain 1.3 M ammonium sulfate by the slow addition of conditioning buffer (50 mM Tris-HCl, 3 M ammonium sulfate, pH 8.0). The 'cut lysate' was incubated at 4°C for 30 min before centrifugation at 30,000 x g for 45 min to remove insoluble precipitate. The clarified sample was loaded onto a 50 mL butyl sepharose HP column (GE Healthcare). Following sample loading, the column was washed using butyl buffer B (50 mM Tris-HCl, 1.3 M ammonium sulfate, pH 8.0). Bound protein was eluted using a gradient to 100% butyl buffer A (50 mM Tris-HCl pH 8.0) over 20 column volumes (1 L). Fractions containing BoNT/F1(0), determined by visual inspection of an SDS-PAGE gel (see section 3.2.1), were pooled and concentrated using a 10MWCO spin concentrator. Concentrated columns were desalted into butyl buffer A using a HiPrep Desalting

Column (GE Healthcare). Desalted protein was loaded onto a 50 ml Q HP column (GE Healthcare) using a flow rate of 10 mL min^{-1}. The column was washed with butyl buffer A and bound protein eluted using a gradient to 25% Q Buffer (50 mM Tris-HCl, 1 M NaCl, pH 8.0) over 2 column volumes (100 ml).

5.2.8 SAXS Data Collection - Batch Mode

Proteins to be analysed by batch mode SAXS were buffer exchanged into phosphate buffered saline (PBS). Each protein sample was diluted 10-fold using PBS and concentrated to 10% of the diluted sample volume using a 50 MWCO spin-concentrator (Millipore). The protein sample buffer exchange was repeated four times. The flow-through was collected at the end of the final buffer exchange and saved for use as the buffer blank SAXS measurement for perfect background subtraction.

5.2.9 SAXS Data Collection - HPLC Mode

Small-angle X-ray Scattering (SAXS) data were collected at beamline B21 of the Diamond Light Source (Oxon, UK), using a PILATUS-6M detector (Dectris, Switzerland). An Agilent 1200 HPLC system on beamline B21 at Diamond Light Source was setup with a Superdex200 column (GE Healthcare) and equilibrated using PBS. Protein samples were concentrated and loaded onto the column at a flow rate of 0.075 mL min^{-1}. Eluted protein was run into a capillary which passed directly through the x-ray beam and data were collected at 15°C using a beam energy of 12.4 keV. Exposures of 1 second were continuously recorded during the run.

5.2.10 SAXS Data Reduction

For batch-mode SAXS, frames corresponding to the buffer blank were averaged and the *Rg* was estimated across the protein exposure using Scatter (Diamond Light Source, UK). Where the estimated *Rg* was stable, frames containing protein scattering signal were averaged and then subtracted by removing the contribution from the averaged buffer sample. For HPLC-mode SAXS the data were processed using software from the ATSAS (version 2.8.3) package ("New developments in the ATSAS program package

for small-angle scattering data analysis." 2012). HPLC-SAXS frames were input to Chromixs (ATSAS) and the peak elution scattering from the HPLC run was subtracted using a stable buffer baseline.

5.2.11 SAXS Data Analysis

To determine the Rg, a Guinier approximation was used by applying a linear fit to log(I) against q^2 within Primus (ATSAS) ("New developments in the ATSAS program package for small-angle scattering data analysis." 2012). The Shapiro-Wilks test for normality was applied to the residuals to ensure a good fit using R (R Core Team, 2018). Theoretical scattering curves from high-resolution X-ray crystallographic structures were generated and fit to experimental data using CRYSOL (ATSAS) with a subtraction constant to account for possible systematic errors.

5.2.12 Molecular Dynamics Simulations

Crystal structures used for molecular dynamics were first manually processed. Ligands and other non-protein molecules present in the crystal structure were stripped from the PDB file using PyMol (v1.8, Schrödinger, LLC) and missing loops were modelled using Rosetta (Simons *et al.*, 1997). All molecular dynamics simulations were performed using GROMACS (Abraham *et al.*, 2015). The proteins were first solvated in a cubic box of pre-equilibrated waters which extended at least 10 Å from any protein atom. The system net charge was neutralised with the addition of Na^+ or Cl^- ions. All simulations were run using the Amber99sb-ILDN (Lindorff-Larsen *et al.*, 2010) force field and the TIP3P water model (Jorgensen *et al.*, 1983; Harrach and Drossel, 2014). Before the main MD runs were performed, the system was energy minimised using the steepest method until the target maximum force ($F_{max} < 1000\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$) was reached. The system was further equilibrated under NVT conditions (constant number of particles, volume, and temperature) for 100 ps with a target temperature of 300 K, followed by NPT conditions (constant number of particles, pressure, and temperature) for another 100 ps each with time steps of 2 fs. All production runs were performed at 300 K, out under NPT conditions for 100 ns with a 2 fs time step.

5.2.13 Molecular Dynamics Analysis

Analysis of each production trajectory was completed using GROMACS (Abraham *et al.*, 2015). Each trajectory was modified by centering the protein molecule within the box to correct for positional jumps when the protein left one side of the box boundary. The corrected trajectories were used to calculate the RMSD, RMSF and Rg. Data were averaged across trajectories for the RMSF and average Rg calculations using R (R Core Team, 2018).

5.3 Results and Discussion

5.3.1 Molecular Dynamics

The crystal structure of BoNT/E1 reveal a compact domain organisation which may have implications for the speed of translocation during intoxication. Molecular dynamics was used to further investigate whether the crystal structure is a stable conformation in solution and is not due to crystal packing. Computational simulation of BoNT/E1 in a hydrated box was completed in order to assess the stability of the 'closed' crystal structure conformation. The crystal structure (PDB ID: 3FFZ Swaminathan and Eswaramoorthy (2000)) was initially manually processed to remove all non-protein atoms with exception to the LC active-site zinc, and missing amino acid side-chains were added. A small disordered region of the belt (S_{460}–N_{466}), for which there were no atoms in the crystal structure, was built *ab initio* using Rosetta (Simons *et al.*, 1997). The molecular dynamics system was created and the protein placed in a cubic box, solvated with a 3-point water model which extended at least 10 Å from any protein atom. Chloride ions were also added to ensure the protein charges remained neutral. The structure was energy minimised until the system reached a stable equilibrium after ~ 3 ns (Figure 5.3).

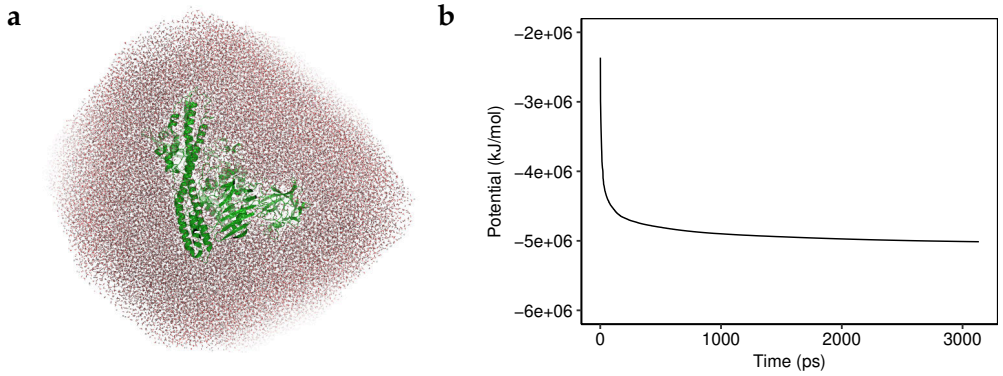


Figure 5.3: BoNT/E MD Energy Minimisation
The BoNT/E crystal structure was energy minimised within the molecular dynamics system until the potential energy of the protein was stable.

The crystal structure of BoNT/A1 was also subjected to molecular dynamic simulation for comparison against our BoNT/E1 data. The structure was manually processed to add missing amino-acid side chains and the molecular dynamics system was setup in the same manner as BoNT/E. The system was energy minimised for ~ 4 ns until the system reached a stable equilibrium (Figure 5.4).

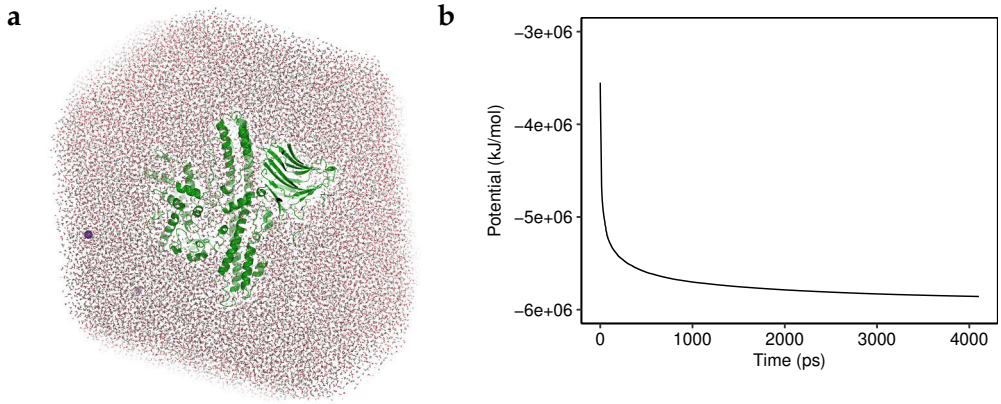


Figure 5.4: BoNT/A MD Energy Minimisation

The BoNT/A crystal structure was energy minimised within the molecular dynamics system until the potential energy of the protein was stable.

Each energy-minimised structure was then used as a starting point to produce three independent molecular dynamics simulations (trajectories) of 100 ns for analysis. The simulations progressed at 5.4 ns d^{-1} on 20 threads of a local server containing an Intel® Xeon® processor E5-2630v4. The total running time was over 110 days. For both BoNT/A1 and BoNT/E1, the three trajectories were reasonably stable across the full time period with an maximum RMSD from the starting model of 5.2 Å (Figure 5.5).

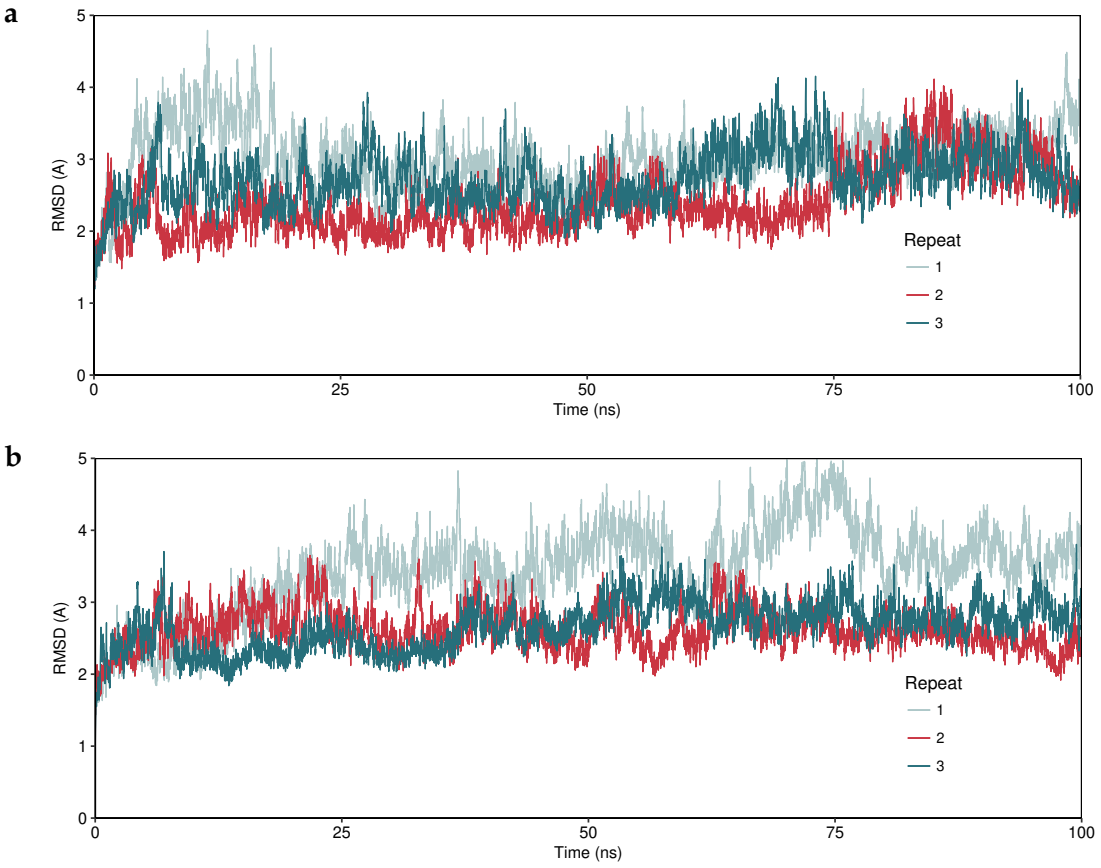


Figure 5.5: Molecular Dynamics RMSD

The RMSD from the starting energy-minimised structure of **(a)** BoNT/A1 and **(b)** BoNT/E1 were calculated over the course of three independent 100 ns molecular dynamics simulations.

To investigate if there is a difference in mobility difference between the domains of BoNT/A1 and BoNT/E1, the root mean squared fluctuation of each C_α atom was calculated (Figure 5.6). Comparing two different structures can be challenging, especially when they contain different protein chain lengths. In the case of BoNTs, a general comparison is possible due to the distinct domain boundaries. Overall, BoNT/A1 displays a larger average RMSF value (1.52 Å) than BoNT/E1 (1.25 Å), alluding to a higher degree of flexibility within the structure. The most striking differences are observed in the H_{CC}, for which BoNT/E1 has a lower RMSF value likely due to the additional structural contacts made between the LC and H_{CN}. By far the largest fluctuation observed for either protein was a disordered loop (BoNT/A1: Y_{752}–I_{765}; BoNT/E1: Y_{726}–I_{743};) at one extreme end of the H_N domain where the RMSF > 4 Å. Very little is known about the BoNT H_N domain but it is likely to form a pore which would require a rearrangement of the domain structure.

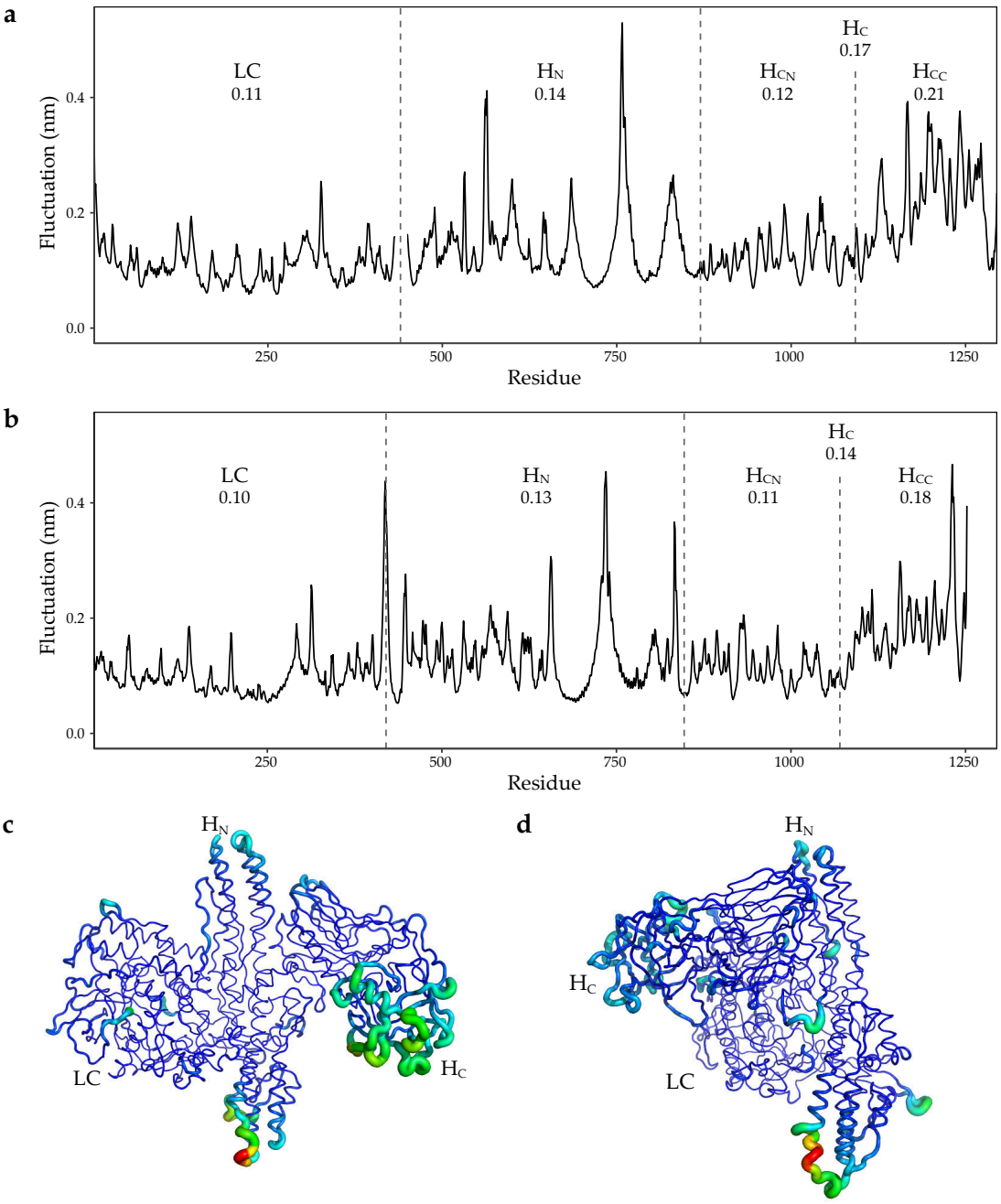


Figure 5.6: Molecular Dynamics RMSF

The average root mean squared fluctuation (RMSF) for each residue of **(a)** BoNT/A1 and **(b)** BoNT/E1 over the course of three independent 100 ns molecular dynamics simulations. Domain boundaries are shown as dashed vertical lines and the average RMSF for each domain is given. The same RMSF values are represented as a cartoon putty on the crystal structures of **(c)** BoNT/A1 and **(d)** BoNT/E1.

The trajectory of each run was used to generate clusters which each represent a highly sampled conformation of BoNT/E. None of the conformations generated represent an opening of the structure, similar to that of BoNT/A or BoNT/B. The distance between two residues, one within the H_C domain and another within the LC domain, was calculated for each time point within each of the trajectories. The distance remains stable, indicating the H_C does not move, even transiently, with respect to the LC.

5.3.2 Purification of Full-length BoNTs

Full-length BoNTs were produced for SAXS analysis. Each protein construct contained two mutations within the HExxH zinc-binding motif to disable catalytic endopeptidase activity and ensure that the protein was non-toxic – these are termed endonegative BoNTs. Such mutations have been shown to have no large effects on the structure of the protein (Gu *et al.*, 2012).

BoNT/A1

First an endonegative BoNT/A1 was expressed and purified. The protein was produced without a purification tag and required multiple purification steps including hydrophobic interaction chromatography and ion exchange chromatography (Figure 5.7. The final purified product was of a high purity (Figure 5.8).

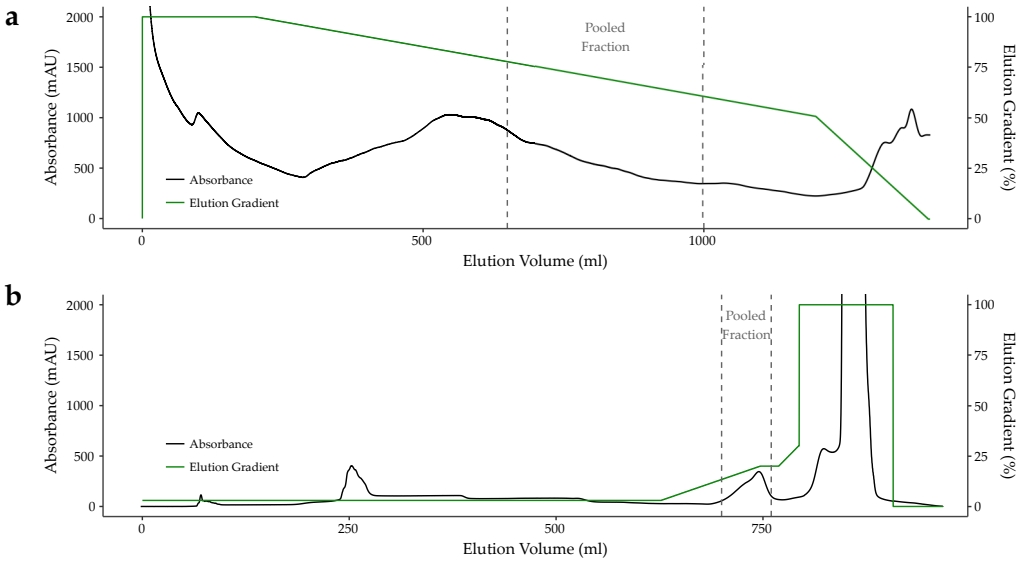


Figure 5.7: Purification of BoNT/A1(0)

(a) HIC purification of BoNT/A1(0) failed to fully separate the protein. **(b)** Further purification by IEC removed the contaminants yielding purified BoNT/A1(0).

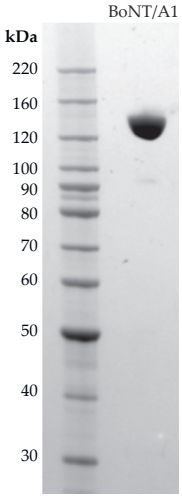


Figure 5.8: BoNT/A1 Purification SDS-PAGE

4%–12% gradient SDS-PAGE of the final purified sample of endonegative BoNT/A1.

BoNT/F1

BoNT/F1 was purified for SAXS analysis using a two step purification procedure 5.9. A small contaminant was still present after the second purification column 5.10 but this should have no effect on the final SAXS data collection as the BoNT was to be further purified.

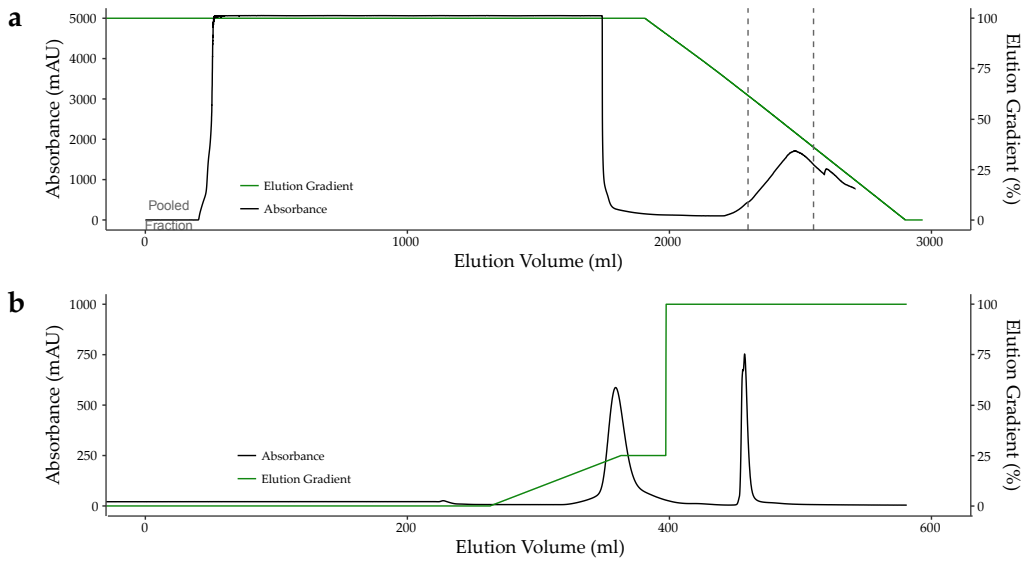


Figure 5.9: Purification of BoNT/F1

(a) Protein was eluted from the a hydrophobic interaction chromatography column using a decreasing gradient of ammonium sulphate. **(a)** BoNT/F1 was further purified by ion exchange chromatogaphy.

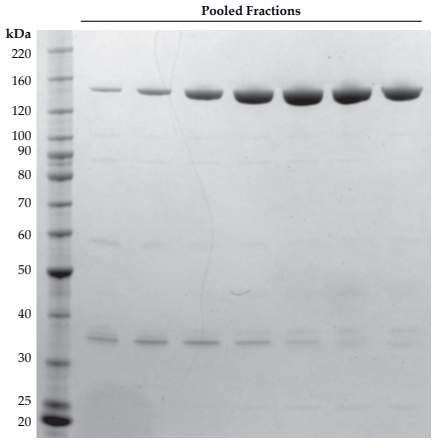


Figure 5.10: BoNT/F1 Purification SDS-PAGE

4%–12% gradient SDS-PAGE of the final purified sample of endonegative BoNT/F1.

BoNT/FA

BoNT/FA was also purified for SAXS studies. The BoNT/FA construct contained a 10x histidine-tag at the N-terminus which was used for purification by IMAC followed by size exclusion chromatography (Figure 5.11).

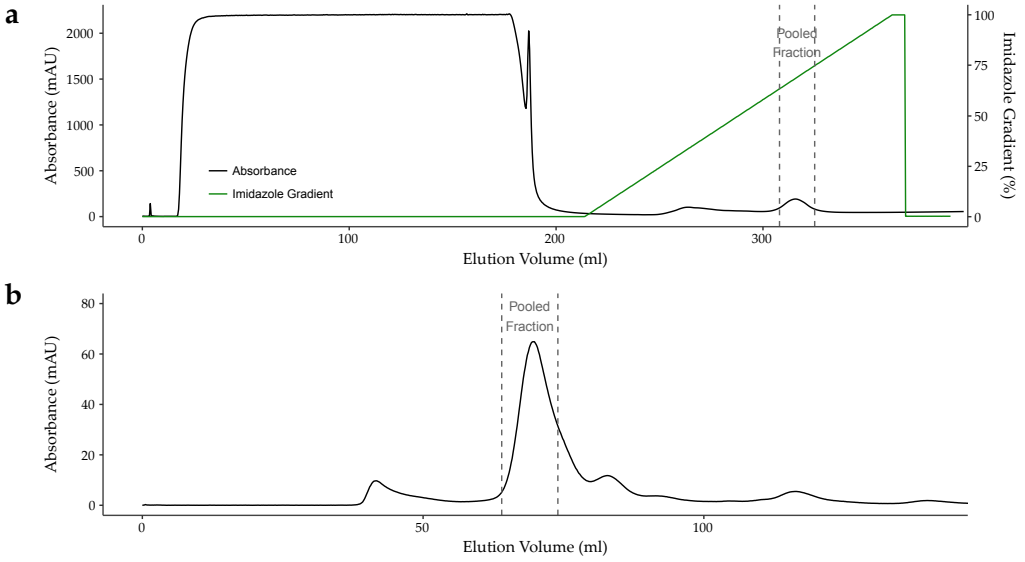


Figure 5.11: Purification of BoNT/FA

(a) IMAC purification of BoNT/FA resulted in a large single peak of eluted protein. Clarified *E. coli* lysate was loaded onto a HisTrap column and the column washed before a gradient elution. **(b)** Size exclusion chromatography removed small residual contaminants and imidazole from the IMAC purificaiton.

In addition to SAXS, crystallisation of BoNT/FA(0) was attempted by screening using the crystallisation screening method described in Chapter 3. Initial concentration of BoNT/FA(0) resulted in the appearance of thin plate crystals at concentrations above $2\,\mathrm{mg\,mL^{-1}}$. The crystals were mounted and X-ray diffraction data collection was attempted at room temperature on a home source. No diffraction was observed and instead the crystals were crushed for use as a seed stock for crystallisation screening. Screening resulted in no hits for BoNT/FA(0).

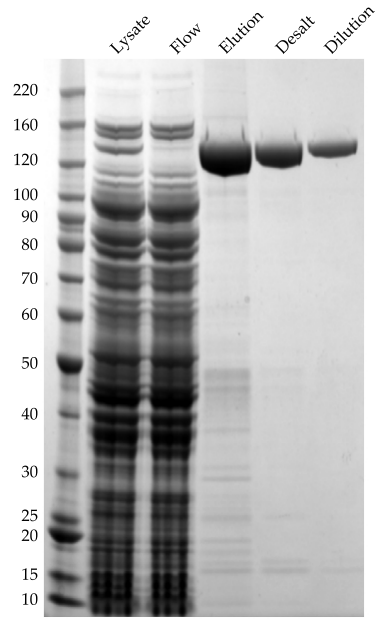


Figure 5.12: BoNT/FA(0) Purification
SDS-PAGE of each step from the purification of BoNT/FA(0). Protein from lysate was captured on a HisTrap column (GE Healthcare) and the elution desalted. Final purified protein was diluted to $0.5\,\mathrm{mg\,mL^{-1}}$.

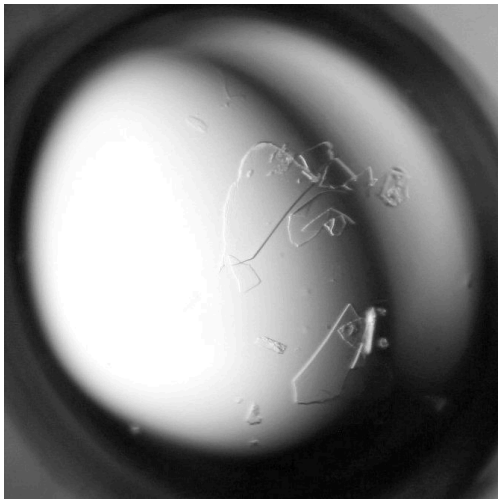


Figure 5.13: BoNT/FA(0) Crystal Morphology
A 1 µL drop containing thin plate crystals of BoNT/FA(0) following concentration of the purified protein.

5.3.3 BoNT/A1(0) and BoNT/E1(0) SAXS

Data Collection and Processing

The X-ray crystal structures of BoNT/A, BoNT/B and BoNT/E have previously been determined revealing a unique domain rearrangement of BoNT/E compared with an open 'butterfly' arrangement of BoNT/A and /B. To determine whether the BoNT/A, BoNT/B and BoNT/E crystal structures are representative of the native conformation in solution, we collected small-angle x-ray scattering (SAXS) data for each. SAXS data was collected on beamline B21 at Diamond Light Source in either batch- or hplc-mode. Initial batch-mode data collected for BoNT/E1(0) showed strong indications of protein aggregation. At low q the scattering intensity should plateau toward $q = 0$, but instead for the batch-mode sample, the intensity increased toward $q = 0$ (Figure 5.14b). To improve the data quality, data was collected by another method, SEC-SAXS, for BoNT/E1(0). Consecutive x-ray exposures were taken for the duration of the SEC run and a buffer sample from the start of the run was subtracted (Figure 5.14a). To ensure the largest peak on the SEC run was a single species, the Rg was calculated across the peak. Buffer-subtracted data for the SEC-SAXS run was significantly better than data collected for the batch-mode run and protein aggregates were not present (Figure 5.14b).

Due to the nature of SEC-SAXS, the protein sample BoNT/E1 was less concentrated than at the start of the run and so the data has a low signal-to-noise and increased noise at higher q (Figure 5.16). HPLC-mode SAXS data were also collected for BoNT/A1, BoNT/B1 and BoNT/F1 (Figure 5.15). The X-ray scattering intensity was recorded over the duration of the HPLC run and can therefore be plotted to display the purification chromatogram. Immediately clear is the difference in elution profiles between BoNT/E and the other BoNTs (/A, /B and /F), suggesting it is a smaller or more compact molecule. The molecular weight of BoNT/E1 (143.8 kDa) is smaller than BoNT/A (149.5 kDa), /B (150.8 kDa), and /F (147.1 kDa) and therefore this may also cause a shift toward the right of the size exclusion chromatography peak. The Rg value was calculated across the peak of each HPLC run and the data for each protein was background subtracted. Guinier plots were used to find the linear scattering region at low q in order to determine an approximate radius of gyration (Rg) (Figure 5.16).

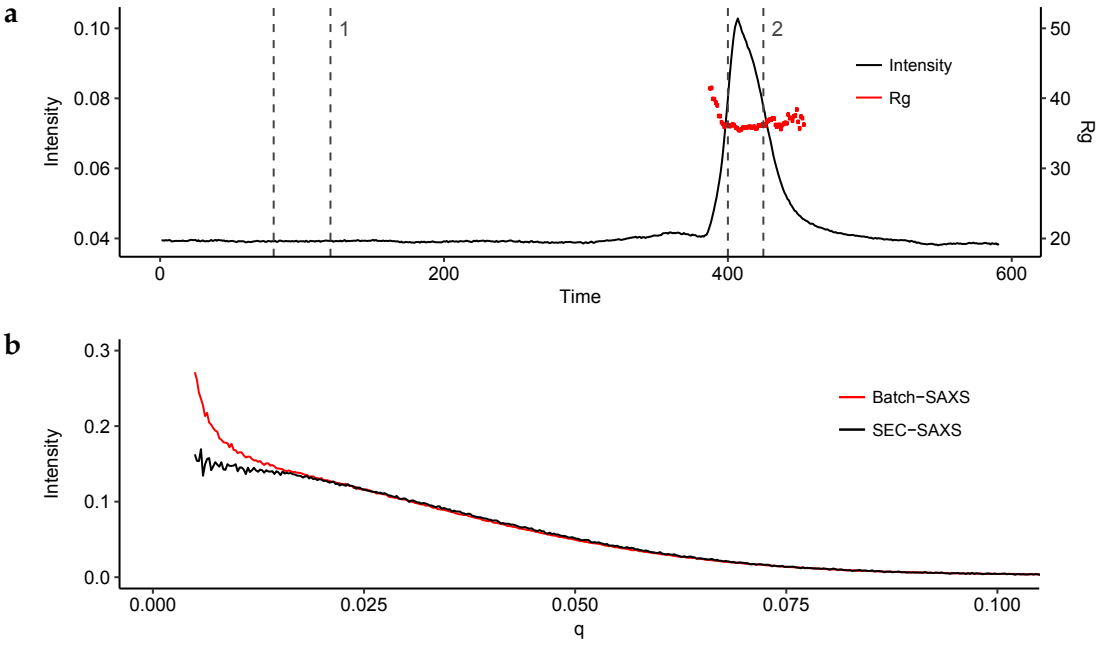


Figure 5.14: BoNT/E1 SEC-SAXS

(a) Intensity of X-ray scattering over time as the BoNT/E1(0) preparation passed through a Superdex200 (GE Healthcare) size-exclusion chromatography column. A series of buffer exposures (area 1) were integrated, averaged, and subtracted from integrated protein exposures (area 2) to produce the final subtracted dataset. The *Rg* (red points) was calculated for each exposure along the elution peak. **(b)** Buffer-subtracted intensities for X-ray scattering data collected using two methods, Batch-SAXS and SEC-SAXS, with the former displaying particle aggregation.

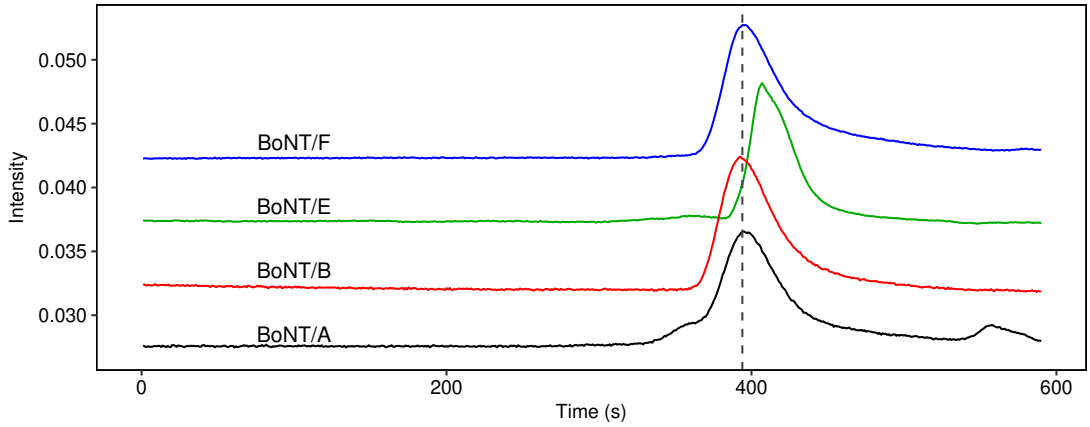


Figure 5.15: HPLC-SAXS Elution Profiles The HPLC-SAXS elution profiles were inferred from measured X-ray scattering during the purification run. Intensities have been scaled and translated to aid visual appearance. The elution peaks for BoNT/A, /B, and /F are approximately indicated by the vertical dashed line.

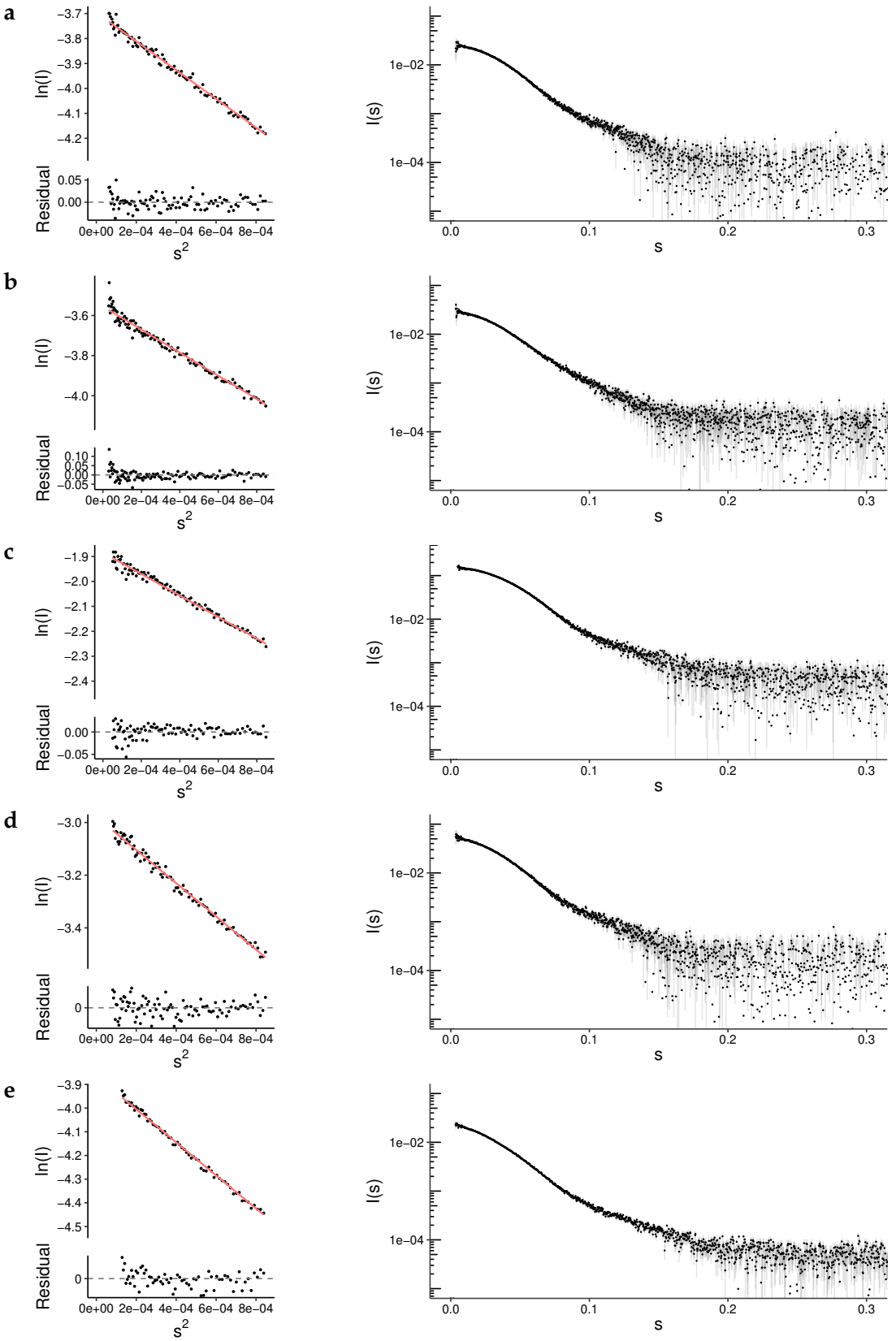


Figure 5.16: BoNT SAXS Data and Guinier Approximation

Guinier fit and residuals (Left) and recorded SAXS data (Right) for BoNT/A1 **(a)**, BoNT/B1 **(b)**, BoNT/E1 **(c)**, BoNT/f1 **(d)**, and BoNT/FA **(e)**.

The crystal structures of BoNT/A and BoNT/B display an 'open' domain organisation similar to each other, therefore we would expect them to have a similar *Rg* in solution. Indeed, the experimental *Rg* of BoNT/A1 and BoNT/B1 was 41.37 Å and 40.79 Å respectively (Table 5.1). An approximate *Rg* was also calculated from full-length crystal structures (Rg_{PDB}). The BoNT/E1 crystal structure displays a 'closed' domain organisation, if the crystal structure was similar to the solution structure then a smaller *Rg* would be expected. The experimental *Rg* of BoNT/E1 was 35.95 Å, smaller than both BoNT/A and BoNT/B, and near to the calculated *Rg* from its crystal structure (Table 5.1). These initial data indicate strongly that BoNT/E1 adopts a more compact structure than BoNT/A1 in solution. The data collected for for BoNT/A1, /B1, and /E1 can be used as standards for the investigation into the solution fold of other BoNTs. It is feasible that collection of SAXS data for any BoNT will enable us to confidently determine if the BoNT is globular in nature (such as BoNT/E1) or in the 'open butterfly' fold (such as BoNT/A1), or anywhere in between.

Table 5.1: Small-angle X-ray Scattering Parameters

	BoNT/A	BoNT/B	BoNT/E	BoNT/F	BoNT/FA
I0	0.025	0.028	0.15	0.05	0.021
Rg (Å)	41.37	40.79	35.95	43.01	45.62
Rg_{PDB} (Å)	39.31[a]	40.59[b]	35.23[c]	–	–
Rg_{MD} (Å)	39.72	–	35.81	–	–
D_{max}	138.64	139.84	115.11	133.65	NC
Model Fit χ^2					
BoNT/A[a]	**1.064**	**1.073**	6.089	**1.174**	2.969
BoNT/B[b]	**1.178**	**0.949**	6.183	**1.221**	2.404
BoNT/E[c]	2.666	3.689	**0.968**	3.226	9.085

[a] PDB: 3BTA. [b] PDB: 1S0E. [c] PDB: 3FFZ. †

Calculated Theoretical Scattering

The existing crystal structures of BoNT/A1, /B1, and /E1 can be used to generate theoretical SAXS data and then compared directly to the experimental scattering data. Theoretical data were calculated using Crysol and compared to the experimental data for each BoNT, χ^2 values were produced detailing their agreement (Table 5.1). To emphasise the agreement between the experimental scattering and theoretical scattering curves, the data can be presented as a normalised, dimensionless-Kratky plot, where the scattering intensities are scaled and presented as $(qR_g)^2 I(q)/I(0)$ versus qR_g (Figure 5.17). A dimensionless-Kratky plot emphasises the low q region of data and allows examination of the overall shape of the molecule. This normalised plot gives an indication of protein conformation, showing peak maxima at $qR_g = \sqrt{3}$ and $(qR_g)^2 I(q)/I(0) = 1.104$ for globular particles independent of the concentration and size of the sample. Peak deviations suggest higher levels of protein flexibility are present.

The theoretical scattering data for both BoNT/A and BoNT/B fit experimental scattering data for either BoNT/A or BoNT/B very well ($\chi^2 \approx 1$, Table 5.1), suggesting both proteins adopt a similar conformation in solution. In addition, the theoretical scattering data for BoNT/E matches its respective experimental data very well too. Neither theoretical data for BoNT/A nor BoNT/B fit the BoNT/E experimental data or *vice versa*. The degree to which the crystal structures match their respective solution scattering data is impressive and clearly confirms the existence to two distinct conformations across these BoNT serotypes.

Domain Organisation of BoNT/F1 and BoNT/FA

No full-length crystal structures exist of BoNT/F1 or BoNT/FA despite our attempts. SAXS-data was collected for BoNT/FA using batch-mode but unfortunately some aggregation was observed at low q. SAXS data collected for BoNT/F1 was good and appeared to have no-aggregation issues. The experimental Rg for BoNT/F1 was 43.01 Å, much higher than the Rg of BoNT/E1 and closer to BoNT/A1 and /B1 (Table 5.1). The theoretical scattering from the full-length crystal structures were also compared to the scattering of BoNT/F1 and the χ^2 values indicate a strong similarity

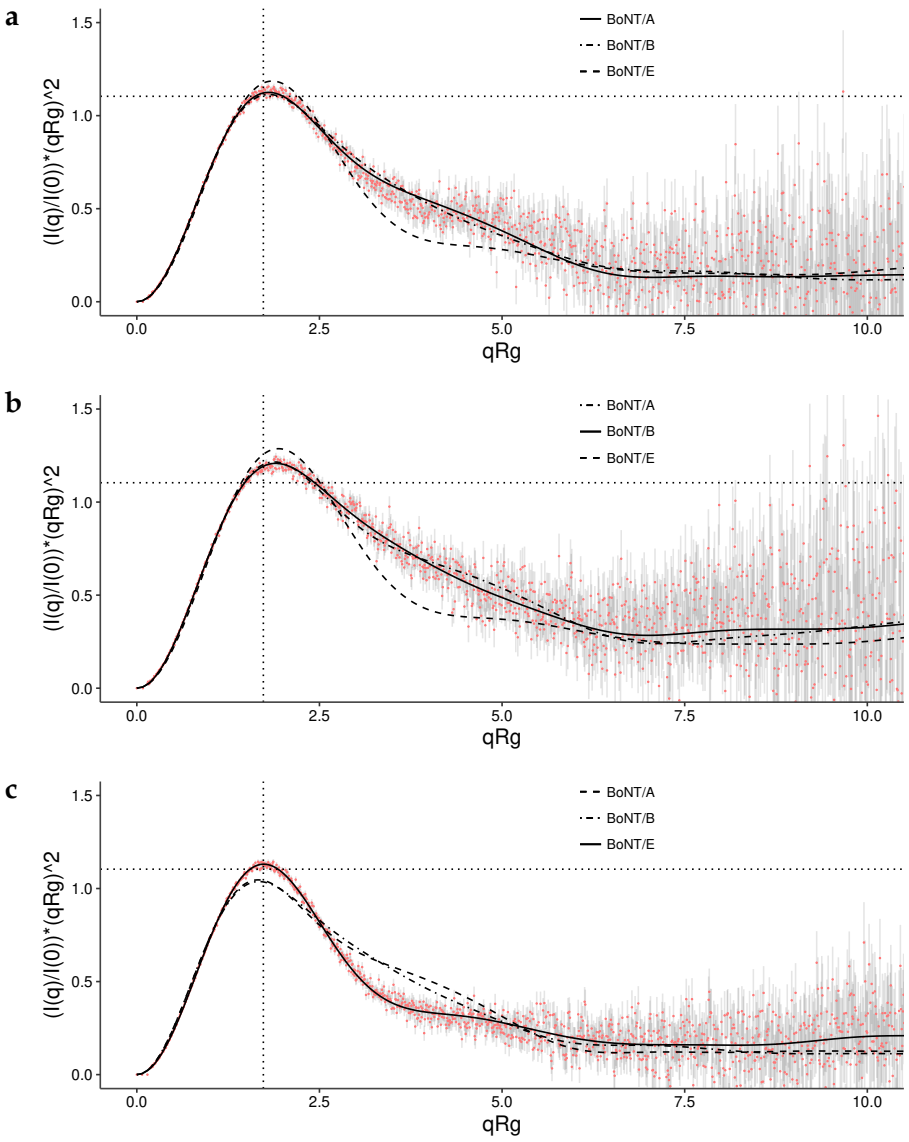


Figure 5.17: SAXS Normalised Kratky Plot

A normalised Kratky plot of the SAXS from samples of **(a)** BoNT/A1, **(b)** BoNT/B1, and **(c)** BoNT/E1. Solid and dashed lines indicate the calculated scattering values from the crystal structures of BoNT/A, /B, and /E.

to both the BoNT/A1 and /B1 structures. We suggest that BoNT/F1 in solution has an 'open' domain organisation as opposed to the 'closed' domain organisation present in BoNT/E1.

Unfortunately, due to the aggregation present in BoNT/FA we cannot be confident in any of the calculated parameters. However, despite poor χ^2 values for each crystal structure fit to the SAXS data (Table 5.1), there does appear to be a indication that BoNT/FA likely also has an 'open' domain organisation. Improved SAXS data, which would likely be possible through the use of HPLC-SAXS, is needed to remove the present aggregation and provide increased confidence.

5.4 Conclusions

Understanding translocation is one of the few unknowns remaining in the botulinum neurotoxin field. The crystal structure of BoNT/E revealed a novel arrangement of the domains in the molecule when compared with other BoNTs and it has been suggested that this may allow the protein to translocation faster into the cytoplasm. While the BoNT/E structure had this conformation *in crystallo*, it was not known whether the similar conformation existed in solution or whether it was a result of a crystallogaphic artefact. We first investigated the stability of the BoNT/A1 and BoNT/E1 crystal structures *in silico* using molecular dynamics simulations. After multiple repeats of 100 ns trajectories for each protein, the stability of their domain organisation was analysed. The domain organisation of both proteins was found stable over the course of 100 ns. The simulations suggest that the H_C domain of BoNT/E may be structurally more stable than BoNT/A due to the increased hydrogen bonding with the rest of the molecule. No domain reorganisation was present during the simulation period.

To further investigate the stability of the BoNT/E1 structure, we turned to SAXS – a technique for investigating protein structure in solution. We first determined the solution state of BoNT/E1 using SAXS and compared with other BoNTs with a known crystal structure. Our work confirmed that BoNT/E does indeed have the same structure in solution as in the crystal as determined by a comparison of the scattering data with calculated scattering from the crystal structure (χ^2 0.968). The SAXS data for BoNT/A1 and /B1 also correlate with their respective crystal structures very well,

confirming that there are two unique domain organisations present across these three serotypes. Further, we propose that SAXS is an excellent technique for investigating the domain organisation of other full-length BoNTs from natural or engineered sources in the future. We assessed the structure of a BoNT with an unknown crystal structure, BoNT/F1 and suggest that the solution structure is the same as BoNT/A and BoNT/B. Investigation of another protein, BoNT/FA, was less successful due to aggregation issue which could have been potentially avoided through the use of HPLC-mode SAXS. In our experience, all full-length BoNTs display some aggregation at concentrations above $1\,\mathrm{mg\,mL^{-1}}$ and therefore we recommend always utilising HPLC-mode SAXS if available ensure the best possible data collection.

CHAPTER 6

Creation of a Botulinum Neurotoxin Database

6.1 Introduction

Botulinum neurotoxins (BoNTs) are categorised into serotypes based on a system of anti-sera recognition (serotyping) and then further divided in subtypes. The first *Clostridum botulinum* genome was sequenced in 2007 and since then many more unique BoNT variants have been identified (Sebaihia *et al.*, 2007; Hill and Smith, 2013). While the method of serotyping BoNTs remains, newly identified proteins can now be accurately characterised into a specific subtype through sequence comparison. The recent improvements in DNA sequencing has meant it is not uncommon to sequence entire genomes and this has lead to many novel BoNT discoveries (Shendure *et al.*, 2017; Brunt *et al.*, 2018; Zhang *et al.*, 2017b).

The current nomenclature holds that if a protein sequence has greater than 2.5% sequence variation to its closed neighbour then it must be a new subtype (Peck *et al.*, 2017). Unfortunately, this system results in multiple unique proteins sharing the same subtype name – for example, there are fourteen sequences identified as BoNT/B2 which differ by up to 2.17% (28 amino acid residues of 1291 total). It has been established that BoNT subtypes may display differences of intoxication however it is less well understood if highly similar sequences of the same subtype may also display intoxication differences. There are certainly some examples where relatively a single residue changes can have a drastic effect. For example, the binding domain (H_C) of BoNT/FA has low solubility which can be increased significantly through a $M_{1148}R$ mutation (Yao *et al.*, 2017). The variations found between BoNTs of the same subtype may hold useful information for future protein design. Another problem with some BoNT literature currently is that it is common to refer to the toxin used for research by simply its subtype or even serotype name. Here, it is difficult to determine the exact protein sequence used, and in many cases impossible without contacting the authors directly. On occasion the bacterial strain which produces the toxin of interest, may be named which can aid identification of the sequence but searching existing databases with only a strain can be very difficult. Further confusing the BoNT nomenclature, are bacterial strains which produce multiple toxin types (Franciosa *et al.*, 2009).

A recent review, authored by many in the BoNT field, has proposed the use of a database of protein sequences for the purpose of categorising and identifying

novel BoNTs and their variants (Peck *et al.*, 2017). Presented in this chapter is the development of a new BoNT database which may fulfil this need and encourage responsible reporting of specific BoNT sequences in future.

6.2 Methods

6.2.1 Automated Protein Sequence Retrieval

Known BoNT protein sequences are used for BLAST searches against the National Center for Biotechnology Information (NCBI) non-redundant protein database using the BLAST search binaries (Camacho *et al.*, 2009) wrapped in a custom Python script. Returned protein sequences and associated publications are retrieved through programmatic access to the Entrez system (NCBI, 2018). Associated UniprotKB accession codes are retrieved through programmatic access to the Uniprot application programming interface (API) (Nightingale *et al.*, 2017). Non-duplicate sequences are added to the BoNTbase database using a custom Python script which has access to the internal MySQL database. Sequence searches are performed automatically on a weekly basis but only if the NCBI database contains an update.

6.2.2 Development of BoNTbase

The BoNTbase website was built manually using a combination of PHP, Python, HTML, and CSS languages. The full source code will be made available to an researchers wishing to view it.

6.3 Results and Discussion

6.3.1 Automated Sequence Retrieval

The process of searching protein sequence databases, such as those at the NCBI or UniProtKB, can be time consuming and tedious when looking for every protein within a specific family. In many cases, proteins found within these databases lack some or most of the appropriate annotations, making it difficult to determine the protein origin, subtype or serotype. To address these problems first we decided to produce our own directory of BoNT protein sequences. A custom workflow was written in Python, termed 'BoNTblast', that allowed automation of the search process and included validation of the returned sequences (Figure 6.1). The first step involves searching the National Center for Biotechnology Information (NCBI) non-redundant protein sequence database. For the workflow to begin, it was provided with a representative protein sequence from each BoNT serotype (BoNT/A–/G). The initial stages of the workflow utilised BLAST search binaries (Camacho *et al.*, 2009) to query the remote BLAST database hosted at the NCBI. One problem with automating deep database searches such as we are here, is that along with desired protein sequences, there can be many unrelated and unwanted sequences returned. For example, the *bont* gene in *C. botulinum* is normally found immediately adjacent to the *ntnha* gene which encodes a protein of about ~ 20% to 30% identity to its related BoNT (Gu *et al.*, 2012). The NTNHA protein is similar enough that it sometimes appears in BLAST results when querying against a BoNT sequence. Filtering must be carefully planned as too much will result in missed targets. A list of NTNHA protein sequences were manually added to the BoNTblast filter to remove then from the workflow. In future it may be worth incorporating these sequences into the database for added functionality. A total of 151 non-redundant BoNT sequences have been found using the automated sequence retrieval method outlined here (Table 6.1).

6.3.2 Development of a BoNT Database

It was necessary to organise the retrieved data in a manner that was organised and accessible and so the open-source database software, MySQL, was chosen. Non-

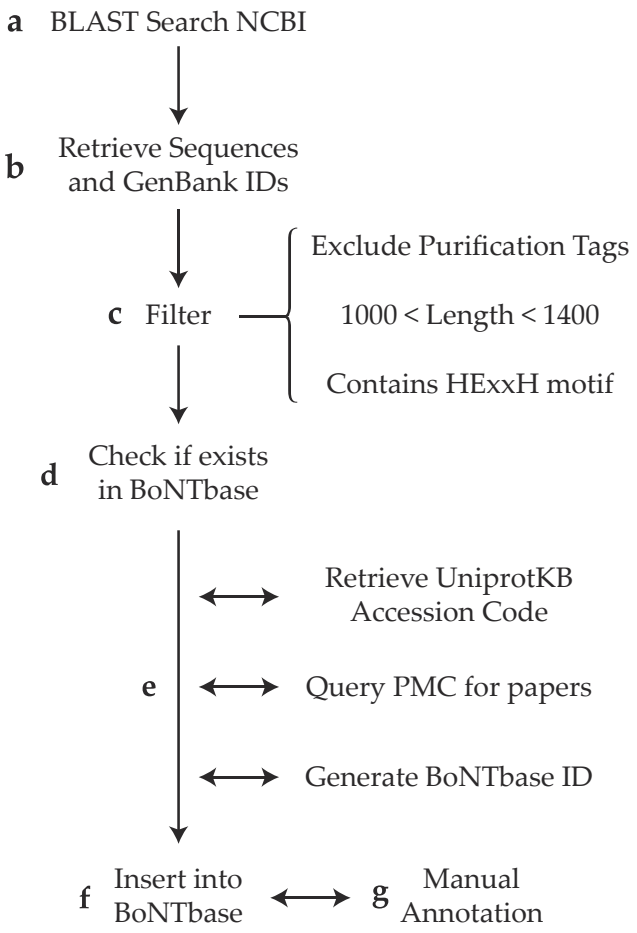


Figure 6.1: Automated BoNTblast sequence retrieval workflow.
(**a**) The NCBI BLAST database is first queried using known BoNT sequences. (**b**) Protein sequences and GenBank IDs are retrieved from BLAST hits. (**c**) Non-redundant sequences are passed through a rule-based filter to remove unlikely BoNT sequences. (**d**) New sequences are finally checked against the BoNTbase database to determine if they already exist. (**e**) New sequences are annotated with their respective UniProtKB IDs (if exists), a list of associated publications from PubMed Central (PMC), and a new unique BoNTbase ID. (**f**) Resulting data and metadata is added to BoNTbase and (**g**) manually annotated to correct any errors.

Table 6.1: Number of identified non-redundant BoNT sequences.
Percentage identities are given for full-length and H_C domain sequence alignments of each BoNT/A subtype[†]. For the H_C alignments, sequences aligned to BoNT/A1 K870 – L1296 were used.

Serotype	No. of Subtypes	Total NR Sequences
BoNT/A	8	29
BoNT/B	8	40
BoNT/C	–	2
BoNT/CD	–	9
BoNT/D	–	5
BoNT/DC	–	5
BoNT/E	12	37
BoNT/F	9	18
BoNT/G	1	2
Others[†]	–	4

[†] BoNT/Wo, BoNT/En, BoNT/HA, BoNT/X.

redundant BoNT sequences discovered using BoNTblast were added to the backend database (MySQL) which allowed fast and simple querying by any other application with programmatic access. Organisation of the database was considered a priority to ensure no protein sequences could be mistaken for one another. There are many ways to refer to each sequence, including by its UniProtKB ID, NCBI reference sequence ID, strain name etc., however these methods are not consistent as each sequence may have multiple identifiers on each service. Peck *et al.* (2017) propose keeping an arbitrary 2.6% amino acid cut-off value when deciding whether or not a BoNT is a new subtype. This may increase confusion surrounding the use of specific BoNT sequences in research and ideally each sequence should be addressable by a single unique identifier which will never change. To address this concern, each sequence was allocated a BoNTbase ID when added to the database. A BoNTbase ID consists of four alphanumeric characters which are case insensitive and assigned at random (*e.g.* C9KA). If a protein sequence should be removed from BoNTbase, the ID would never be reused. This combination gives a total of $36^4 = 1679616$ possibilities and is *unlikely* to ever be expended. All sequences were subjected to manual curation in order to remove unwanted sequences that passed through the filter and assign BoNT subtype names based on associated literature.

6.3.3 Development of an Online User Interface

The BoNTblast workflow described above lead to a database that contained many BoNT sequences and associated metadata, however this data was not easily accessible to users without a knowledge of the SQL language. To address this, a graphical front-end website was developed using modern web languages to ensure all users could have access through any chosen web browser. The website was called BoNTbase – short for the Botulinum NeuroToxin Database – and the domain https://bontbase.org was acquired. The website backend was written in PHP which interfaces with the MySQL database, allowing easy and customisable searching of all available data for the end user. When developing the user interface, the ability to quickly retrieve BoNT sequences using any available information was a priority. On the home page (https://bontbase.org), users are able to search the database by serotype, subtype, strain name, bacterial species, or by protein sequence. When a user selects a protein of interest, they are directed to a page which contains a range of information relevant to that specific sequence (Figure 6.2).

Firstly, the protein sequence is provided in a format which allows the user to directly copy the sequence from the page for their own use. An option is also given that converts the plain sequence directly into FASTA format, containing an appropriate sequence header which is suitable for many online services such as protein sequence alignment using CLUSTAL (Sievers *et al.*, 2011).

It is also useful for users involved in research to have direct and fast access to information about the biochemical properties of the protein. A range of information about the protein sequence is provided to the user including the protein length, molecular weight, estimated absorbance (280 nm) and calculated pI. The absorbance is calculated as the extinction coefficient normalised by molecular weight where the extinction coefficient is estimated from the number of Tyr, Trp, and Cys residues within the protein (Equation 6.1) (Edelhoch, 1967; Gill and von, 1989). The absorbance (of a 0.1% protein solution) can thus be calculated using $\frac{E}{MW}$.

$$E = 1490(\text{No. Tyr}) + 5500(\text{No. Trp}) + 125(\text{No. Cystine}) \tag{6.1}$$

Figure 6.2: BoNTbase protein information.

An example of the layout from a BoNT summary page. **(a)** The subtype name and BoNTbase unique ID. **(b)** The protein sequence. **(c)** Citations for publications which have used this specific protein in their research. **(d)** Information about this page and protein sequence including the molecular weight, length, absorbance (280 nm), and calculated pI. **(e)** A list of all the bacterial strains which produce this protein. **(f)** Links the same sequence in external databases. **(g)** Links to other proteins of the same subtype and their sequence identity.

Where the molar exctinction coefficients of Tyr, Trp and Cystine are 1490 M^{-1} cm^{-1}, 5500 M^{-1} cm^{-1}, and 125 M^{-1} cm^{-1} respectively. The calculated pI is also provided and is calculated using parameters for individual amino acids as described in Bjellqvist *et al.* (1993). In addition to the full-length protein sequence the information includes a list of associated publications that have referred to the protein within their research. The list of publications are automatically annotated, to gather them the BoNTblast workflow searches the PubMed Central (PMC) database using all known identifiers associated with the protein sequence (Figure 6.1). This method is prone to false positives however and publications need to be manually checked for relevance. A full list of publications within our database, sorted by the date of release, is also made available.

When accessing a specific BoNT entry, users are also provided with an up-to-date list of all known bacterial strains from our database (Figure 6.3). It is common for many different bacterial strains to produce a BoNT with the same protein sequence and many publications refer to a specific toxin by its associated bacterial strain identifier. In some cases, knowledge of a bacterial strain is useful as some strains are bivalent meaning they can produce two different BoNTs, for example the strain 111 encodes both BoNT/B and BoNT/X proteins (Zhang *et al.*, 2017b). In an effort to make searching our database as user friendly as possible, we compiled a list of all bacterial strains from the NCBI database which produce a BoNT. A custom script queries the NCBI using previously retrieved accession codes and pulls out information regarding individual strains. Each strain is entered into our database and associated with its respective BoNT.

Strain(s): Hall, ATCC3502, NCTC 13319, ATCC17862, ATCC19397, ATCC449, ATCC25763, VPI 7124, A Hall, ATCC 3502, KF Meyer 33, CDC36956, ATCC 25763, AM1295, ATCC 17862, CDC220726, ATCC 19397, CDC220727, 32A, CDC36923, CDC36955, CDC36924, CDC 21547, McClung 844, Nan 2006, Cam2A, 69A, ATCC 449, DFPST0029, Prevot 697B, Hall A, Prevot 910, 73A, KF Meyer 126, Prevot Dewping, Hall A-hyper, BS-A, Allergan-Hall A

Figure 6.3: BoNTbase Strain Information
An example of the bacterial strains associated with a BoNT/A1 subtype. When accessing information regarding a specific BoNT, the user is provided a list of bacterial strains which all encode the selected protein.

6.3.4 User Contributions

Users of BoNTbase are encouraged to upload relevant publications through the web interface if they are aware of missing entries (Figure 6.4). In an effort to keep the list of publications in our database organised, users are simply asked for a Pubmed ID or a digital object identifier (DOI). The website will use the provided information and automatically retrieve all publication information, including author names, the title, and the journal, from the NCBI Pubmed database. The new publication entry is manually curated before it it fully integrated into the database.

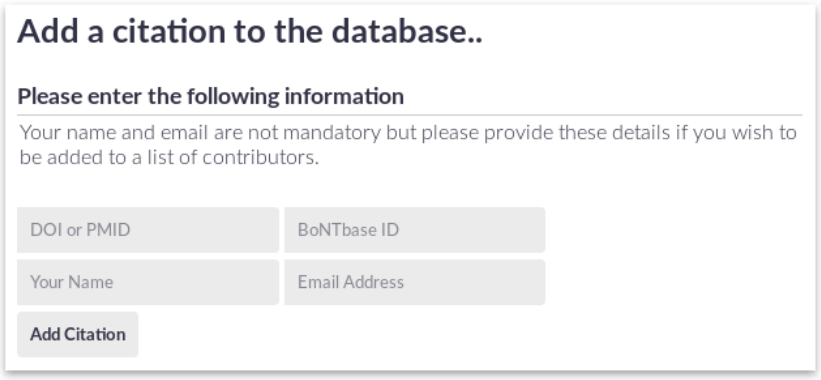


Figure 6.4: BoNTbase User Contribution Form
Users of BoNTbase may provide publication information that is relevant to specific BoNT entries in BoNTbase through an online form requiring only the Pubmed ID or DOI.

6.3.5 Identification of a BoNT-like protein

In addition to collecting all available known BoNT sequences, the BoNTblast workflow also managed to uncover a protein sequence that had the hallmarks of a BoNT but was previously unknown to the BoNT community. The protein was found in a *Enterococcus fecalis* genome (strain 3G1 DIV0629), uploaded to the NCBI in May 2017 by members of The Broad Institute Genomic Center for Infectious Diseases (USA). Annotation of the genome revealed its source to be cow faeces. Our workflow had uncovered a hypothetical protein of 149 kDa in size which had only 31% sequence identity to

its nearest BoNT (BoNT/X) and another 143 kDa protein which we identified as a likely NTNHA protein. Initial analysis of the BoNT-like protein revealed the classic HExxH motif, found in all other BoNTs. This protein was subsequently co-discovered by other groups and was determined to be a BoNT-like protein (Zhang *et al.*, 2018; Brunt *et al.*, 2018). Despite not intending to identify novel proteins the methods used in BoNTbase have proven that this is possible and the identification of further related proteins already in protein databases will be explored in future work.

6.4 Conclusions

BoNTbase is a new protein database that aims to contain every known BoNT sequence and display them in a user friendly and easily searchable way. What was initially designed as a local database to aid research, has now grown into an online resource which is used regularly by scientific groups around the world. The methods used to identify all BoNT sequences could easily be transferred to other areas of research and may help others to also identify novel protein variants that are hidden in the wealth information provided by whole-genome sequencing projects. We hope that this database proves useful to researchers in the BoNT community and feel it has addressed some concerns as reviewed by Peck *et al.* (2017). While BoNTbase is an independent database, it sources all information from a range of other publicly accessible databases and as such, will always remain an open resource for all.

CHAPTER 7

General Discussion and Future Work

Botulinum neurotoxins are a marvel of protein evolution, every step of their mechanism throughout intoxication is tailored specifically to the host in order to cause efficient and specific inhibition on SNARE-mediated exocytosis (Mansfield and Doxey, 2018). The main causative agent of botulism is the bacteria *C. botulinum* which is found natively in soil all over the world. The sequencing of bacterial genomes from patients suffering the symptoms of botulism or from soil samples has led to over 150 unique BoNT variants having been identified (Sebaihia *et al.*, 2007; Mansfield and Doxey, 2018). These BoNTs are categorised into serotypes and then subtypes based on protein sequence differences. Despite the availability of a large number of variants, up until recently commercially available BoNTs for therapeutic use have only focussed on two BoNT subtypes, A1 and B1 (Kane *et al.*, 2015; Fonfria *et al.*, 2018). Other subtypes have been found to have different intoxication properties and they should be investigated as a potential source of novel therapeutics. The work presented in this thesis has resulted in the determination of nine protein crystal structures of BoNT H_C domains to very high resolution (Figure 7.1). In doing so the precise atomic locations and organisation of the BoNT receptor-binding sites have all been mapped and are available to other researchers. Each displays high levels of structural similarity to all of the other H_C domains reported in this thesis (Table 7.1) indicating their different characteristics are the result of subtle changes to sequence and structure alone.

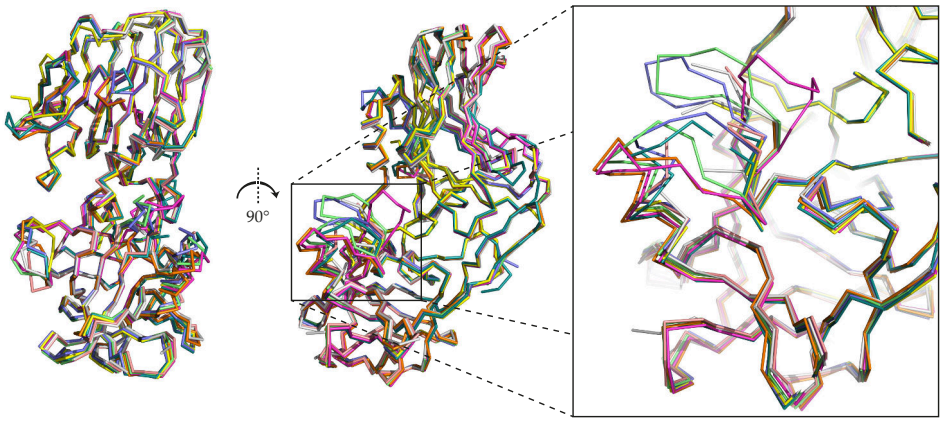


Figure 7.1: Overlay of Reported H_C Crystal Structures
All structures described in Chapter 3 superposed and displayed as ribbons. The ganglioside binding site is enlarged to highlight variation in the region.

Table 7.1: H_C domain RMSD values

Each H_C crystal structure from chapter 3 was superposed using PyMOL (Schrödinger, LLC, 2015) and the RMSD values are reported below.

	A1(1)	**A1(2)**	**A2**	**A3**	**A4**	**A5**	**A6**	**FA**
A1(1)	–							
A1(2)	0.554	–						
A2	0.511	0.887	–					
A3	0.637	0.969	0.331	–				
A4	0.496	0.400	0.637	0.750	–			
A5	0.439	0.524	0.677	0.739	0.494	–		
A6	0.356	0.563	0.447	0.569	0.387	0.390	–	
FA	0.584	0.423	0.586	0.899	0.501	0.581	0.614	–

The high structural conservation found within the structures published here, also suggest that the BoNT H_C may tolerate attempts to modify receptor binding sites. The targeting of different ganglioside types for example, would change neuronal cell specificity which may prove to have therapeutic benefits. Our crystal structures of BoNT/A H_C domains provide a starting point for investigating residues involved in binding to both the ganglioside and SV2-protein receptor. The structures can serve as a basis for future protein engineering and will aid in guiding specific mutations which may improve binding affinity. Further structures of the H_C domains in complex with receptors should be investigated and, as seen in our reported structure of H_C/A3 in complex with the ganglioside GD1a, can give information about the impact of natural variation on receptor-binding. A key variation in the H_C/A3 ganglioside binding site, which also occurs in H_C/A2 and /A6, is tolerated and does not affect the location of the bound ganglioside despite the loss of hydrogen bonding to the Sia5 moiety of GD1a. The generation of a mutant to replicate the H_C/A1 GBS is worth pursuing to determine if this variation effects ganglioside binding *in vivo*.

The role of a disulfide bond within the H_C domain also raises as to its importance *in vivo*. Some of our determined crystal structures contained the C_{1235} to C_{1280} covalent bond while others did not, despite their high conservation among all BoNT subtypes the role is unknown. These residues may play a part in stabilising the protein fold which could potentially have implications on ganglioside binding due to the location being near to the ganglioside binding site. We determined the crystal structures from to crystal forms of H_C/A1 in the same crystallisation condition from the same batch of protein with each structure displaying either the disulfide bond or not. This

raises questions as to whether both forms are produced during protein expression or if the subsequent handling and crystallisation causes some of the total protein to become oxidised or reduced. Analysis such as mass spectrometry may be able to determine the oxidation state. This may have major implications for the production of recombinant BoNTs, especially those produced outside of the native host in *E. coli* for example, of which there now exist some in clinical trials (Fonfria *et al.*, 2018). The high resolution crystal structures of H_C/A5 and H_C/A6 also pose further peculiarities with respect to this disulfide. In neither structure is a C–C bond present, instead a neighbouring residue Lys forms a covalent bond with C_{1280}, in the case of H_C/A5 this is through a methylene-bridge while in H_C/A6 the N_ζ of the Lys is bound directly to the S_γ of the Cys. Whether or not these bonds are as a direct result of the conditions experienced *in crystallo* including the X-ray radiation is unknown. However, low-dose X-ray diffraction data still indicates the presence of the bond. Determining if these covalent bonds are present in native BoNTs would be quite a challenge and different approaches including NMR and mass spectrometry have been listed as possible methods.

In addition to determining high resolution crystal structures, we also investigated binding of the protein receptor, SV2C, to the H_C/A1 to /A6 subtypes. Purification of glycosylated SV2C was successful but the binding assay used had some limitations and was unable to generate binding parameters for H_C/A4 or H_C/A5. Other methods to determine the binding affinity of BoNTs to their receptors should be explored and would generate valuable information which would aid any future protein design. Due to the multiple protein-receptors and different types of ganlgiosides involved it may be advantageous to explore the characterisation of receptor-binding to whole cells. SV2 is a glycoprotein when expressed in mammalian cells. The glycosylation is heterogeneous and BoNTs may have a preference for a specific glycosylation pattern – this is yet to be determined. The specific glycosylation pattern in HEK293T cells (as used for SV2 production in this thesis) may differ from the pattern seen in the motor neuron and this would give rise to limitations of using any recombinant expression system outside of neuronal cells.

Of all the BoNT serotypes, only three (BoNT/A, /B, and /E) have their full-length crystal structures determined. The crystal structure of BoNT/E revealed a compact domain organisation in contrast to the 'butterfly' structure observed in BoNT/A and

BoNT/B. We explored the use of small-angle X-ray scattering (SAXS) to determine the structure of BoNTs in solution and examine whether the crystal structure of BoNT/E is similar. Our work has confirmed the existence of two unique domain organisations across BoNT serotypes as previously identified through X-ray crystallography (Lacy *et al.*, 1998; Kumaran *et al.*, 2009; Swaminathan and Eswaramoorthy, 2000) and provided a robust method for determining the solution structure of other BoNTs. Using easily determined parameters directly from the measured SAXS data and from comparisons to existing crystal structures, it is immediately obvious which domain organisation is present. The HPLC-mode of SAXS measurements is ideal for the investigation of full-length BoNTs as they are prone to low levels of aggregation at relatively low protein concentration which would otherwise interfere with SAXS measurements. This method can be extended to collect SAXS data a various pH levels, simulating the pH gradient found within the endosome during BoNT intoxication, and could potentially give rise to new data toward understanding BoNT translocation. Very little is understood about the mechanism of BoNT translocation and any method that can provide some insight into the process would be of benefit.

We have also presented BoNTbase – a botulinum neurotoxin database which aims to contain all known BoNT protein sequences and provide easy access to their associated publications. The database is programmed to automatically update itself and add new BoNT sequences as soon as they are present in other databases such as Uniprot. In doing so we hope the resource helps researchers to identity all BoNTs and ensure accurate reporting with regard to the specific sequence used. Currently the database holds only information relevant to BoNTs and the bacterial strains which produce them but this could be extended in the future to include other related molecules. The methods and scripts used to retrieve all known sequences can be easily transferred to other fields and will be shared freely. Hopefully to aid researchers in their analyses and potentially discover novel related proteins relevant to them.

References

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, and Lindahl E (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1–2**, 19–25.

Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, and Adams PD (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.

Assmann G, Brehm W, and Diederichs K (2016). Identification of rogue datasets in serial crystallography. *J. Appl. Crystallogr.*, **49**, 1021–1028.

Atassi MZ, Taruishi M, Naqvi M, Steward LE, and Aoki KR (2014). Synaptotagmin II and gangliosides bind independently with botulinum neurotoxin B but each restrains the other. *Protein J.*, **33**, 278–288.

Ayyar BV, Aoki KR, and Atassi MZ (2015). The C-terminal heavy-chain domain of botulinum neurotoxin A is not the only site that binds neurons, as the N-terminal heavy-chain domain also plays a very active role in toxin-cell binding and interactions. *Infect. and Immun.*, **83**, 1465–1476.

Barash JR and Arnon SS (2014). A novel strain of *Clostridium botulinum* that produces type B and type H botulinum toxins. *J. Infect. Dis.*, **209**, 183–191.

Benecke R (2012). Clinical relevance of botulinum toxin immunogenicity. *BioDrugs*, **26**, e1–9.

Benoit RM, Frey D, Hilbert M, Kevenaar JT, Wieser MM, Stirnimann CU, McMillan D, Ceska T, Lebon F, Jaussi R, Steinmetz MO, Schertler GFX, Hoogenraad CC, Capitani G, and Kammerer RA (2014). Structural basis for recognition of synaptic vesicle protein 2C by botulinum neurotoxin A. *Nature*, **505**, 108–111.

Benoit RM, Schärer MA, Wieser MM, Li X, Frey D, and Kammerer RA (2017). Crystal structure of the BoNT/A2 receptor-binding domain in complex with the luminal domain of its neuronal receptor SV2C. *Sci. Rep.*, **7**, 43588.

Benson MA, Fu Z, Kim JJP, and Baldwin MR (2011). Unique ganglioside recognition strategies for clostridial neurotoxins. *J. Biol. Chem.*, **286**, 34015–34022.

Berliocchi L, Fava E, Leist M, Horvat V, Dinsdale D, Read D, and Nicotera P (2005). Botulinum neurotoxin C initiates two different programs for neurite degeneration and neuronal apoptosis. *J. Cell Biol.*, **168**, 607–618.

Berntsson RPA, Peng L, Dong M, and Stenmark P (2013). Structure of dual receptor binding to botulinum neurotoxin B. *Nat. Comms.*, **4**, 2058.

Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DI, and Owens RJ (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.*, **35**, e45.

Betley MJ, Somers E, and DasGupta BR (1989). Characterization of botulinum type A neurotoxin gene: Delineation of the N-terminal encoding region. *Biochem. Biophys. Res. Commun.*, **162**, 1388–1395.

Bhidayasiri R and Truong DD (2008). Evidence for effectiveness of botulinum toxin for hyperhidrosis. *J. Neural. Transm.*, **115**, 641–645.

Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S, and Hochstrasser D (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, **14**, 1023–1031.

Blanchet CE and Svergun DI (2013). Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. *Annu. Rev. Phys. Chem.*, **64**, 37–54.

Blasi J, Chapman ER, Link E, Binz T, Yamasaki S, Camilli PD, Südhof TC, Niemann, and Jahn R (1993a). Botulinum neurotoxin A selectively cleaves the synaptic protein SNAP-25. *Nature*, **365**, 160–163.

Blasi J, Chapman ER, Yamasaki S, Binz T, Niemann H, and Jahn R (1993b). Botulinum neurotoxin C1 blocks neurotransmitter release by means of cleaving HPC-1/syntaxin. *EMBO*, **12**, 4821–4828.

Bragg WH and Bragg WL (1913). The Reflexion of X-rays by Crystals. *Proc. R. Soc. Lond.*, **88**, 428–438.

Breidenbach MA and Brünger AT (2004). Substrate recognition strategy for botulinum neurotoxin serotype A. *Nature*, **432**, 925–929.

Broennimann C, Eikenberry EF, Henrich B, Horisberger R, Huelsen G, Pohl E, Schmitt B, Schulze-Briese C, Suzuki M, Tomizaki T, Toyokawa H, and Wagner A (2006). The PILATUS 1M detector. *J. Synchrotron Radiat.*, **13**, 120–130.

Brünger AT (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.

Brunger AT and Rummel A (2009). Receptor and substrate interactions of clostridial neurotoxins. *Toxicon*, **54**, 550–560.

Brunt J, Carter AT, Stringer SC, and Peck MW (2018). Identification of a novel botulinum neurotoxin gene cluster in *Enterococcus*. *FEBS Lett.*, **592**, 310–317.

Cai S, Kukreja R, Shoesmith S, Chang TW, and Singh BR (2006). Botulinum neurotoxin light chain refolds at endosomal pH for its translocation. *Protein J.*, **25**, 455–62.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, and Madden TL (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Carter AT and Peck MW (2015). Genomes, neurotoxins and biology of *Clostridium botulinum* Group I and Group II. *Res. Microbiol.*, **166**, 303–317.

Carugo O and Djinović CK (2005). When X-rays modify the protein structure: radiation damage at work. *Trends Biochem. Sci.*, **30**, 213–219.

Chai Q, Arndt JW, Dong M, Tepp WH, Johnson EA, Chapman ER, and Stevens RC (2006). Structural basis of cell surface receptor recognition by botulinum neurotoxin B. *Nature*, **444**, 1096–1100.

Chalk CH, Benstead TJ, and Keezer M (2014). Medical treatment for botulism. *Cochrane Database Syst. Rev.*, CD008123.

Chapman ER (2002). Synaptotagmin: a Ca^{2+} sensor that triggers exocytosis? *Nat. Rev. Mol. Cell Biol.*, **3**, 498–508.

Chellapandi P and Prisilla A (2017). Structure, function and evolution of *Clostridium botulinum* C2 and C3 toxins: Insight to poultry and veterinary vaccines. *Curr. Protein Pept. Sci.*, **18**, 412–424.

Chellappan G, Kumar R, Santos E, Goyal D, Cai S, and Singh BR (2015). Structural and functional analysis of botulinum neurotoxin subunits for pH-dependent membrane channel formation and translocation. *Biochim. Biophys. Acta*, **1854**, 1510–6.

Chen C, Arai I, Satterfield R, Young SM, and Jonas P (2017). Synaptotagmin 2 is the fast Ca^{2+} sensor at a central inhibitory synapse. *Cell Rep.*, **18**, 723–736.

Chen S and Barbieri JT (2011). Association of botulinum neurotoxin serotype A light chain with plasma membrane-bound SNAP-25. *J. Biol. Chem.*, **286**, 15067–72.

Chen VB, Arendall 3rd WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, and Richardson DC (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.

Coffield JA, Bakry N, Carlson J, Gomella LG, and Simpson LL (1997). In vitro characterization of botulinum toxin types A, C and D action on human tissues: combined electrophysiologic, pharmacologic and molecular biologic approaches. *J. Pharmacol. Exp. Ther.*, **280**, 1489–1498.

Collins MD and East AK (1998). Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins. *J. Appl. Microbiol.*, **84**, 5–17.

Concepcion J, Witte K, Wartchow C, Choo S, Yao D, Persson H, Wei J, Li P, Heidecker B, and Ma W (2009). Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization. *Comb. Chem. High Throughput Screen*, **12**, 791–800.

Dafforn TR (2007). So how do you know you have a macromolecular complex? *Acta Crystallogr. D Biol. Crystallogr.*, **63**, 17–25.

DasGupta BR and Sugiyama H (1972). Role of a protease in natural activation of *Clostridium botulinum* neurotoxin. *Infect. Immun.*, **6**, 587–590.

Davies JR, Hackett GS, Liu SM, and Acharya KR (2018a). High resolution crystal structures of the receptor-binding domain of *Clostridium botulinum* neurotoxin serotypes A and FA. *PeerJ*, **6**, e4552.

Davies JR, Liu SM, and Acharya KR (2018b). Variations in the botulinum neurotoxin binding domain and the potential for novel therapeutics. *Toxins (Basel)*, **10**, E421.

Davies JR, Rees J, Liu SM, and Acharya KR (2018c). High resolution crystal structures of *Clostridium botulinum* neurotoxin A3 and A4 binding domains. *J. Struct. Biol.*, **202**, 113–117.

de Paiva A and Dolly JO (1990). Light chain of botulinum neurotoxin is active in mammalian motor nerve terminals when delivered via liposomes. *FEBS Lett.*, **277**, 171–174.

Dekleva ML and DasGupta BR (1989). Nicking of single chain *Clostridium botulinum* type A neurotoxin by an endogenous protease. *Biochem. Biophys. Res. Commun.*, **162**, 767–772.

Dekleva ML and Dasgupta BR (1990). Purification and characterization of a protease from *Clostridium botulinum* type A that nicks single-chain type A botulinum neurotoxin into the di-chain form. *J. Bacteriol.*, **172**, 2498–2503.

Diederichs K and Karplus P (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.*, **4**, 269–275.

Dong M, Liu H, Tepp WH, Johnson EA, Janz R, and Chapman ER (2008). Glycosylated SV2A and SV2B mediate the entry of botulinum neurotoxin E into neurons. *Mol. Biol. Cell.*, **19**, 5226–5237.

Dong M, Richards DA, Goodnough MC, Tepp WH, Johnson EA, and Chapman ER (2003). Synaptotagmins I and II mediate entry of botulinum neurotoxin B into cells. *J. Cell Biol.*, **162**, 1293–1303.

Dong M, Tepp WH, Liu H, Johnson EA, and Chapman ER (2007). Mechanism of botulinum neurotoxin B and G entry into hippocampal neurons. *J. Cell Biol.*, **179**, 1511–1522.

Dong M, Yeh F, Tepp WH, Dean C, Johnson EA, Janz R, and Chapman ER (2006). SV2 is the protein receptor for botulinum neurotoxin A. *Science*, **312**, 592–596.

Dover N, Barash JR, and Arnon SS (2009). Novel *Clostridium botulinum* toxin gene arrangement with subtype A5 and partial subtype B3 botulinum neurotoxin genes. *J. Clin. Microbiol.*, **47**, 2349–2350.

Dover N, Barash JR, Hill KK, Xie G, and Arnon SS (2014). Molecular characterization of a novel botulinum neurotoxin type H gene. *J. Infect. Dis.*, **209**, 192–202.

Dover N, Barash JR, Hill KK, Davenport KW, Teshima H, Xie G, and Arnon SS (2013). Clostridium botulinum Strain Af84 Contains Three Neurotoxin Gene Clusters: Bont/A2, bont/F4 and bont/F5. *PLoS ONE*, **8**.

Edelhoch H (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, **6**, 1948–54.

Elliott M, Maignel J, Liu SM, Favre-Guilmard C, Mir I, Farrow P, Hornby F, Marlin S, Palan S, Beard M, and Krupp J (2017). Augmentation of VAMP-catalytic activity of botulinum neurotoxin serotype B does not result in increased potency in physiological systems. *PLoS ONE*, **12**, e0185628.

Emsley P, Lohkamp B, Scott WG, and Cowtan K (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.

Engh RA and Huber R (July 1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A*, **47**, 392–400.

Evans HR, Holloway DE, Sutton JM, Ayriss J, Shone CC, and Acharya KR (2004). C3 exoenzyme from *Clostridium botulinum*: structure of a tetragonal crystal form and a reassessment of NAD-induced flexure. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 1502–1505.

Evans P (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 72–82.

Evans PR and Murshudov GN (2013). How good are my data and what is the resolution? *Acta. Crystallogr. D. Biol. Crystallogr.*, **69**, 1204–1214.

Fernández-Salas E, Steward LE, Ho H, Garay PE, Sun SW, Gilmore MA, Ordas JV, Wang J, Francis J, and Aoki KR (2004). Plasma membrane localization signals in the light chain of botulinum neurotoxin. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 3208–13.

Ferreira CMH, Pinto ISS, Soares EV, and Soares HMVM (2015). (Un)suitability of the use of pH buffers in biological, biochemical and environmental studies and their interaction with metal ions – a review. *RSC Adv.*, **5**, 30989–31003.

Fillo S, Giordani F, Anselmo A, Fortunato A, Palozzi AM, De Santis R, Ciammaruconi A, Spagnolo F, Anniballi F, Fiore A, Auricchio B, De Medici D, and Lista F (2015). Draft genome sequence of *Clostridium botulinum* B2 450 strain from wound botulism in a drug user in Italy. *Genome Announc.*, **3**, e00238–15.

Fischer A (2013). Synchronized chaperone function of botulinum neurotoxin domains mediates light chain translocation into neurons. *Curr. Top. Microbiol. Immunol.*, **364**, 115–137.

Fischer A, Garcia-Rodriguez C, Geren I, Lou J, Marks JD, Nakagawa T, and Montal M (2008). Molecular architecture of botulinum neurotoxin E revealed by single particle electron microscopy. *J. Biol. Chem.*, **283**, 3997–4003.

Fischer A and Montal M (2007). Crucial role of the disulfide bridge between botulinum neurotoxin light and heavy chains in protease translocation across membranes. *J. Biol. Chem.*, **282**, 29604–29611.

Fischer A, Sambashivan S, Brünger AT, and Montal M (2012). Beltless translocation domain of botulinum neurotoxin A embodies a minimum ion-conductive channel. *J. Biol. Chem.*, **287**, 1657–1661.

Fonfria E, Maignel J, Lezmi S, Martin V, Splevins A, Shubber S, Kalinichev M, Foster K, Picaut P, and Krupp J (2018). The expanding therapeutic utility of botulinum neurotoxins. *Toxins*, **10**.

Foster KA (2014). *Molecular Aspects of Botulinum Neurotoxin*. Current Topics in Neurotoxicity. Springer New York.

Franciosa G, Maugliani A, Scalfaro C, and Aureli P (2009). Evidence that plasmid-borne botulinum neurotoxin type B genes are widespread among *Clostridium botulinum* serotype B strains. *PLoS ONE*, **4**, e4829.

Fu Z, Chen C, Barbieri JT, Kim JP, and Baldwin MR (2009). Glycosylated SV2 and gangliosides as dual receptors for botulinum neurotoxin serotype F. *Biochemistry*, **48**, 5631–5641.

Garcia-Rodriguez C, Levy R, Arndt JW, Forsyth CM, Razai A, Lou J, Geren I, Stevens RC, and Marks JD (2007). Molecular evolution of antibody cross-reactivity for two subtypes of type A botulinum neurotoxin. *Nature biotechnol.*, **25**, 107–116.

Garcia EA, Trivellin G, Aflorei ED, Powell M, Grieve J, Alusi G, Pobereskin L, Shariati B, Cudlip S, Roncaroli F, Mendoza N, Grossman AB, Harper EA, and Korbonits M (2013). Characterization of SNARE proteins in human pituitary adenomas: targeted secretion inhibitors as a new strategy for the treatment of acromegaly? *J. Clin. Endocrinol. Metab.*, **98**, E1918–E1926.

Gill SC and von Hippel PH (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.*, **182**, 319–26.

Gimenez JA and DasGupta BR (1993). Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, and 18 kD fragments. *J. Prot. Chem.*, **12**, 351–363.

Gonzalez-Escalona N, Thirunavukkarasu N, Singh A, Toro M, Brown EW, Zink D, Rummel A, and Sharma SK (2014). Draft genome sequence of bivalent *Clostridium botulinum* strain

IBCA10-7060, encoding botulinum neurotoxin B and a new FA mosaic type. *Genome Announc.*, **2**.

Gore S, Sanz García E, Hendrickx PMS, Gutmanas A, Westbrook JD, Yang H, Feng Z, Baskaran K, Berrisford JM, Hudson BP, Ikegawa Y, Kobayashi N, Lawson CL, Mading S, Mak L, Mukhopadhyay A, Oldfield TJ, Patwardhan A, Peisach E, Sahni G, Sekharan MR, Sen S, Shao C, Smart O, Ulrich EL, Yamashita R, Quesada M, Young JY, Nakamura H, Markley JL, Berman HM, Burley SK, Velankar S, and Kleywegt GJ (2017). Validation of structures in the Protein Data Bank. *Structure*, **25**, 1916–1927.

Gorrec F (2015). The MORPHEUS II protein crystallization screen. *Acta Crystallogr. F.*, **71**, 831–837.

Gu S, Rumpel S, Zhou J, Strotmeier J, Bigalke H, Perry K, Shoemaker CB, Rummel A, and Jin R (2012). Botulinum neurotoxin is shielded by NTNHA in an interlocked complex. *Science*, **335**, 977–981.

Guinier A (1939). La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys.*, **11**, 161–237.

Gustafsson R, Zhang S, Masuyer G, Dong M, and Stenmark P (2018). Crystal structure of botulinum neurotoxin A2 in complex with the human protein receptor SV2C reveals plasticity in receptor binding. *Toxins (Basel)*, **10**.

Hackett G, Moore K, Burgin D, Hornby F, Gray B, Elliott M, Mir I, and Beard M (2018). Purification and characterization of recombinant botulinum neurotoxin serotype FA, also known as serotype H. *Toxins (Basel)*, **10**, E195.

Hamark C, Berntsson RP, Masuyer G, Henriksson LM, Gustafsson R, Stenmark P, and Widmalm G (2017). Glycans confer specificity to the recognition of ganglioside receptors by botulinum neurotoxin A. *J. Am. Chem. Soc.*, **139**, 218–230.

Harrach M and Drossel B (2014). Structure and dynamics of TIP3P, TIP4P, and TIP5P water near smooth and atomistic walls of different hydroaffinity. *J. Chem. Phys.*, **140**, 174501.

Hatheway CL (1990). Toxigenic clostridia. *Clin. Microbiol. Rev.*, **3**, 66–98.

Hauptman H (1997). Phasing methods for protein crystallography. *Curr. Opin. Struct. Biol.*, **7**, 672–680.

Hauser D, Eklund MW, Kurazono H, Binz T, Niemann H, Gill DM, Boquet P, and Popoff MR (1990). Nucleotide sequence of *Clostridium botulinum* C1 neurotoxin. *Nucleic Acids Res.*, **18**, 4924.

Hermann C (1928). Zur systematischen Strukturtheorie I. Eine neue Raumgruppensymbolik. *Z. Krist.*, **68**, 257–287.

Hexsel C, Hexsel D, Porto MD, Schilling J, and Siega C (2011). Botulinum toxin type A for aging face and aesthetic uses. *Dermatol. Ther.*, **24**, 54–61.

Hill KK and Smith TJ (2013). *Genetic diversity within Clostridium botulinum serotypes, botulinum neurotoxin gene clusters and toxin subtypes*. Springer, pp. 1–20.

Hooper DC and Hirsch MS (2014). Novel *Clostridium botulinum* Toxin and Dual Use Research of Concern Issues. *J. Infect. Dis.*, **209**, 167.

Hug S (2013). Classical molecular dynamics in a nutshell. *Methods Mol. Biol.*, **924**, 127–52.

Jacobson MJ, Lin G, Raphael B, Andreadis J, and Johnson EA (2008). Analysis of neurotoxin cluster genes in *Clostridium botulinum* strains producing botulinum neurotoxin serotype A subtypes. *Appl. Environ. Microbiol.*, **74**, 2778–2786.

Jacques DA and Trewhella J (2010). Small-angle scattering for structural biology–expanding the frontier while avoiding the pitfalls. *Protein Sci.*, **19**, 642–657.

Jin R, Rummel A, Binz T, and Brünger AT (2006). Botulinum neurotoxin B recognizes its protein receptor with high affinity and specificity. *Nature*, **444**, 1092–1095.

Johnson EA (1999). Clostridial toxins as therapeutic agents: benefits of nature's most toxic proteins. *Annu. Rev. Microbiol.*, **53**, 551–575.

Johnson EA (2014). Validity of botulinum neurotoxin serotype H. *J. Infect. Dis.*, **210**, 933–992.

Joosten RP, Long F, Murshudov GN, and Perrakis A (2014). The PDB_REDO server for macromolecular structure model optimization. *IUCrJ*, **1**, 213–220.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, and Klein ML (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

Kamata Y, Kozaki S, Sakaguchi G, Iwamori M, and Nagai Y (1986). Evidence for direct binding of *Clostridium botulinum* type E derivative toxin and its fragments to gangliosides and free fatty acids. *Biochem. Biophys. Res. Commun.*, **140**, 1015–9.

Kamata Y, Kimura Y, Hiroi T, Sakaguchi G, and Kozaki S (1993). Purification and characterization of the ganglioside-binding fragment of *Clostridium botulinum* type E neurotoxin. *Biochem. Biophys. Res. Commun.*, **1156**, 213–218.

Kammerer RA and Benoit RM (2014). Botulinum neurotoxins: new questions arising from structural biology. *Trends Biochem. Sci.*, **39**, 517–526.

Kane CD, Nuss JE, and Bavari S (2015). Novel therapeutic uses and formulations of botulinum neurotoxins: a patent review (2012-2014). *Expert Opin. Ther. Pat.*, **25**, 675–690.

Karalewitz AP, Kroken AR, Fu Z, Baldwin MR, Kim JP, and Barbieri JT (2010). Identification of a unique ganglioside binding loop within botulinum neurotoxins C and D-SA. *Biochemistry*, **49**, 8117–8126.

Karalewitz A, Fu Z, Baldwin MR, Kim JP, and Barbieri JT (2012). Botulinum neurotoxin serotype C associates with dual ganglioside receptors to facilitate cell entry. *J. Biol. Chem.*, **287**, 40806–40816.

Karplus PA and Diederichs K (2015). Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.*, **34**, 60–68.

Keller JE, Cai F, and Neale EA (2004). Uptake of botulinum neurotoxin into cultured neurons. *Biochemistry*, **43**, 526–532.

Kelley LA, Mezulis S, Yates CM, Wass MN, and Sternberg MJ (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.

Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wychoff H, and Phillips DC (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, **181**, 662–626.

Kerner J (1817). Vergiftung durch verdorbene Würste. *Tübinger Blätter für Naturwissenschaften und Arzneykunde*, **3**, 1–25.

Kerscher M, Roll S, Becker A, and Wigger-Alberti W (2012). Comparison of the spread of three botulinum toxin type A preparations. *Arch. Dermatol. Res.*, **304**, 155–161.

Kessler KR, Skutta M, and Benecke R (1999). Long-term treatment of cervical dystonia with botulinum toxin A: efficacy, safety, and antibody frequency. German Dystonia Study Group. *J. Neurol.*, **246**, 265–274.

Kohda T, Ihara H, Seto Y, Tsutsuki H, Mukamoto M, and Kozaki S (2007). Differential contribution of the residues in C-terminal half of the heavy chain of botulinum neurotoxin type B to its binding to the ganglioside GT1b and the synaptotagmin 2/GT1b complex. *Microb. Pathog.*, **42**, 72–79.

Koriazova LK and Montal M (2003). Translocation of botulinum neurotoxin light chain protease through the heavy chain channel. *Nat. Struct. Biol.*, **10**, 13–8.

Kroken AR, Blum FC, Zuverink M, and Barbieri JT (2017). Entry of botulinum neurotoxin subtypes A1 and A2 into neurons. *Infect. Immun.*, **85**, e00795–16.

Kroken AR, Karalewitz AP, Fu Z, Baldwin MR, Kim JJ, and Barbieri JT (2011a). Unique ganglioside binding by botulinum neurotoxins C and D-SA. *FEBS J.*, **278**, 4486–96.

Kroken AR, Karalewitz AP, Fu Z, Kim JJ, and Barbieri JT (2011b). Novel ganglioside-mediated entry of botulinum neurotoxin serotype D into neurons. *J. Biol. Chem.*, **286**, 26828–37.

Kull S, Schulz KM, Weisemann J, Kirchner S, Schreiber T, Bollenbach A, Dabrowski PW, Nitsche A, Kalb SR, Dorner MB, Barr JR, Rummel A, and Dorner BG (2015). Isolation and functional characterization of the novel *Clostridium botulinum* neurotoxin A8 subtype. *PLoS ONE*, **10**, e0116381.

Kumaran D, Eswaramoorthy S, Furey W, Navaza J, Sax M, and Swaminathan S (2009). Domain organization in *Clostridium botulinum* neurotoxin yype E is unique: its implication in faster translocation. *J. Mol. Biol.*, **386**, 233–245.

Lacy DB, Tepp W, Cohen AC, DasGupta BR, and Stevens RC (1998). Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nat. Struct. Biol.*, **5**, 898–902.

Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis J, Dror R, and Shaw D (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950–1958.

Lindström M, Nevas M, Kurki J, Sauna-aho R, Latvala-Kiesilä A, Pölönen I, and Korkeala H (2004). Type C botulism due to toxic feed affecting 52,000 farmed foxes and minks in Finland. *J. Clin. Microbiol.*, **42**, 4718–25.

Low BW, Chen CCH, Berger JE, Singman L, and Pletcher JF (1966). Studies of insulin crystals at low temperatures: effects on mosaic character and radiation sensitivity. *Proc. Natl. Acad. Sci.*, **56**, 1746–1750.

Luft JR, Newman J, and Snell EH (2014). Crystallization screening: the influence of history on current practice. *Acta Crystallogr. F Struct. Biol. Commun.*, **70**, 835–853.

Luquez C, Raphael BH, and Maslanka SE (2009). Neurotoxin gene clusters in *Clostridium botulinum* type Ab strains. *Appl. Environ. Microbiol.*, **75**, 6094–6101.

Mahrhold S, Bergström T, Stern D, Dorner BG, Åstot C, and Rummel A (2016). Only the complex N559-glycan in the synaptic vesicle glycoprotein 2C mediates high affinity binding to botulinum neurotoxin serotype A1. *Biochem. J.*, **473**, 2645–2654.

Mahrhold S, Rummel A, Bigalke H, Davletov B, and Binz T (2006). The synaptic vesicle protein 2C mediates the uptake of botulinum neurotoxin A into phrenic nerves. *FEBS letters*, **580**, 2011–2014.

Mansfield MJ, Adams JB, and Doxey AC (2015). Botulinum neurotoxin homologs in non-Clostridium species. *FEBS Lett.*, **589**, 342–348.

Mansfield MJ and Doxey AC (2018). Genomic insights into the evolution and ecology of botulinum neurotoxins. *Pathog. Dis.*, **76**.

Masuyer G, Chaddock JA, Foster KA, and Acharya KR (2014). Engineered botulinum neurotoxins as new therapeutics. *Annu. Rev. Pharmacol. Toxicol.*, **54**, 27–51.

Masuyer G, Davies JR, Moore K, Chaddock JA, and Acharya KR (2015). Structural analysis of *Clostridium botulinum* neurotoxin type D as a platform for the development of targeted secretion inhibitors. *Sci. Rep.*, **5**, 13397.

Masuyer G, Zhang S, Barkho S, Shen Y, Henriksson L, Košenina S, Dong M, and Stenmark P (2018). Structural characterisation of the catalytic domain of botulinum neurotoxin X - high activity and unique substrate specificity. *Sci. Rep.*, **8**, 4518.

Mauguin C (1931). Sur le symbolisme des groupes de répétition ou de symétrie des assemblages cristallins. *Z. Krist.*, **76**, 542–558.

Mazuet C, Dano J, Popoff MR, Créminon C, and Volland H (2010). Characterization of botulinum neurotoxin type A neutralizing monoclonal antibodies and influence of their half-lives on therapeutic activity. *PLoS ONE*, **5**, e12416.

Mazuet C, Ezan E, Volland H, Popoff MR, and Becher F (2012). Toxin detection in patients' sera by mass spectrometry during two outbreaks of type A botulism in France. *J. Clin. Microbiol.*, **50**, 4091–4094. arXiv: `mdl-22993181`.

McCoy AJ (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.*, **63**, 32–41.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ (2007). Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.

McNicholas S, Potterton E, Wilson KS, and Noble ME (2011). Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 386–394.

Montecucco C and Molgó J (2005). Botulinal neurotoxins: revival of an old killer. *Curr Opin Pharmacol*, **5**, 274–279.

Montecucco C (1986). How do tetanus and botulinum toxins bind to neuronal membranes? *Trends Biochem. Sci.*, **11**, 314–317.

Morgans CW, Kensel-Hammes P, Hurley JB, Burton K, Idzerda R, McKnight GS, and Bajjalieh SM (2009). Loss of the synaptic vesicle protein SV2B results in reduced neurotransmission and altered synaptic vesicle protein expression in the retina. *PLoS ONE*, **4**, e5230.

Moriishi K, Koura M, Fujii N, Fujinaga Y, Inoue K, Syuto B, and Oguma K (1996). Molecular cloning of the gene encoding the mosaic neurotoxin, composed of parts of botulinum neurotoxin types C1 and D, and PCR detection of this gene from *Clostridium botulinum* type C organisms. *Appl. Environ. Microbiol.*, **62**, 662–667.

Mueller M, Wang M, and Schulze-Bries C (2012). Optimal fine ϕ-slicing for single-photon-counting pixel detectors. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 42–56.

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, and Vagin AA (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta. Crystallogr. D. Biol. Crystallogr.*, **67**, 355–367.

Murshudov GN, Vagin AA, and Dodson EJ (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.

Nakamura K, Kohda T, Umeda K, Yamamoto H, Mukamoto M, and Kozaki S (2010). Characterization of the D/C mosaic neurotoxin produced by *Clostridium botulinum* associated with bovine botulism in Japan. *Vet. Microbiol.*, **140**, 147–154.

NCBI (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **46**, D8–D13.

New developments in the ATSAS program package for small-angle scattering data analysis. (2012). *J. Appl. Crystallogr.*, **45**, 342–350.

Newman J, Egan D, Walter TS, Meged R, Berry I, Ben Jelloul M, Sussman JL, Stuart DI, and Perrakis A (2005). Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 1426–31.

Neylon C (2008). Small angle neutron and X-ray scattering in structural biology: recent examples from the literature. *Eur. Biophys. J.*, **37**, 531–541.

Nightingale A, Antunes R, Alpi E, Bursteinas B, Gonzales L, Liu W, Luo J, Qi G, Turner E, and Martin M (2017). The Proteins API: accessing key integrated protein and genome information. *Nucleic Acids Res.*, **45**, W539–W544.

Nishiki T, Kamata Y, Nemoto Y, Omori A, Ito T, Takahashi M, and Kozaki S (1994). Identification of protein receptor for *Clostridium botulinum* type B neurotoxin in rat brain synaptosomes. *J. Biol. Chem.*, **269**, 10498–10503.

Nishiki T, Tokuyama Y, Kamata Y, Nemoto Y, Yoshida A, Sekiguchi M, Takahashi M, and Kozaki S (1996). Binding of botulinum type B neurotoxin to Chinese hamster ovary cells transfected with rat synaptotagmin II cDNA. *Neurosci. Lett.*, **208**, 105–8.

Ogasawara J, Kamata Y, Sakaguchi G, and Kozaki S (1991). Properties of a protease-sensitive acceptor component in mouse brain synaptosomes for *Clostridium botulinum* type B neurotoxin. *FEMS Microbiol. Lett.*, **79**, 351–356.

Overly CC, Lee KD, Berthiaume E, and Hollenbeck PJ (1995). Quantitative measurement of intraorganelle pH in the endosomal-lysosomal pathway in neurons by using ratiometric imaging with pyranine. *Proc. Natl. Acad. Sci.*, **92**, 3156–60.

Pal D and Chakrabarti P (2001). Non-hydrogen bond interactions involving the methionine sulfur atom. *J. Biomol. Struct. Dyn.*, **19**, 115–128.

Parkhurst JM, Thorn A, Vollmar M, Winter G, Waterman DG, Fuentes-Montero L, Gildea RJ, Murshudov GN, and Evans G (2017). Background modelling of diffraction data in the presence of ice rings. *IUCrJ*, **4**, 626–638.

Patel K, Cai S, and Singh BR (2014). Current strategies for designing antidotes against botulinum neurotoxins. *Expert Opin. Drug. Discov.*, **9**, 319–333.

Patterson AL (1935). A direct method for the determination of the components of interatomic distances in crystals. *Zeitschrift für Kristallographie*, 517.

Peck MW, Smith TJ, Anniballi F, Austin J, Bano L, Bradshaw M, Cuervo P, Cheng L, Derman Y, Dorner BG, Fisher A, Hill K, Kalb SR, Korkeala H, Lindström M, Lista F, Lúquez C, Mazuet C, Pirazzini M, Popoff M, Rossetto O, Rummel A, Sesardic D, Singh BR, and Stringer SC (2017). Historical perspectives and guidelines for botulinum neurotoxin subtype nomenclature. *Toxins (Basel)*, **9**, E38.

Pellett S, Bradshaw M, Tepp WH, Pier CL, Whitemarsh RCM, Chen C, Barbieri JT, and Johnson EA (2018). The light chain defines the duration of action of botulinum toxin serotype A subtypes. *MBio*, **9**, e00089–18.

Pellett S, Tepp WH, Whitemarsh RC, Bradshaw M, and Johnson EA (2015). In vivo onset and duration of action varies for botulinum neurotoxin A subtypes 1-5. *Toxicon*, **107**, 37–42.

Peng L, Berntsson RP, Tepp WH, Pitkin RM, Johnson EA, Stenmark P, and Dong M (2012). Botulinum neurotoxin D-C uses synaptotagmin I and II as receptors, and human synaptotagmin II is not an effective receptor for type B, D-C and G toxins. *J. Cell. Sci.*, **125**, 3233–42.

Peng L, Liu H, Ruan H, Tepp WH, Stoothoff WH, Brown RH, Johnson EA, Yao WD, Zhang SC, and Dong M (2013). Cytotoxicity of botulinum neurotoxins reveals a direct role of syntaxin 1 and SNAP-25 in neuron survival. *Nat. Commun.*, **4**, 1472.

Peng L, Tepp WH, Johnson EA, and Dong M (2011). Botulinum neurotoxin D uses synaptic vesicle protein SV2 and gangliosides as receptors. *PLoS Pathog.*, **7**, e1002008.

Petrova T, Ginell S, Mitschler A, Kim Y, Lunin VY, Joachimiak G, Cousido-Siah A, Hazemann I, Podjarny A, Lazarski K, and Joachimiak A (2010). X-ray-induced deterioration of disulfide bridges at atomic resolution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 1075–1091.

Pier CL, Chen C, Tepp WH, Lin G, Janda KD, Barbieri JT, Pellett S, and Johnson EA (2011). Botulinum neurotoxin subtype A2 enters neuronal cells faster than subtype A1. *FEBS Lett.*, **585**, 199–206.

Pirazzini M, Azarnia Tehran D, Leka O, Zanetti G, Rossetto O, and Montecucco C (2016). On the translocation of botulinum and tetanus neurotoxins across the membrane of acidic intracellular compartments. *Biochim. Biophys. Acta*, **1858**, 467–474.

Pirazzini M, Azarnia Tehran D, Zanetti G, Megighian A, Scorzeto M, Fillo S, Shone CC, Binz T, Rossetto O, Lista F, and Montecucco C (2014). Thioredoxin and its reductase are present on synaptic vesicles, and their inhibition prevents the paralysis induced by botulinum neurotoxins. *Cell Rep.*, **8**, 1870–1878.

Pirazzini M, Rossetto O, Bolognese P, Shone CC, and Montecucco C (2011). Double anchorage to the membrane and intact inter-chain disulfide bond are required for the low pH induced entry of tetanus and botulinum neurotoxins into neurons. *Cell. Microbiol.*, **13**, 1731–1743.

Prabakaran S, Tepp W, and DasGupta BR (2001). Botulinum neurotoxin types B and E: purification, limited proteolysis by endoproteinase Glu-C and pepsin, and comparison of their identified cleaved sites relative to the three-dimensional structure of type A neurotoxin. *Toxicon*, **39**, 1515–1531.

Przedpelski A, Tepp WH, Kroken AR, Fu Z, Kim JJP, Johnson EA, and Barbieri JT (2013). Enhancing the protective immune response against botulism. *Infect. Immun.*, **81**, 2638–2644.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Ramachandran GN, Ramakrishnan C, and Sasisekharan V (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, **7**, 95–99.

Read RJ, Adams PD, Arendall WB, Brünger AT, Emsley P, Joosten RP, Kleywegt GJ, Krissinel EB, Lütteke T, Otwinowski Z, Perrakis A, Richardson JS, Sheffler WH, Smith JL, Tickle IJ, Vriend G, and Zwart PH (2011). A new generation of crystallographic validation tools for the protein data bank. *Structure*, **19**, 1395–1412.

Relman DA (2014). "Inconvenient truths" in the pursuit of scientific knowledge and public health. *J. Infect. Dis.*, **209**, 170–172.

Robert X and Gouet P (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, **42**, 320–4.

Rossetto O, Pirazzini M, and Montecucco C (2014). Botulinum neurotoxins: genetic, structural and mechanistic insights. *Nat. Rev. Microbiol.*, **12**, 535–549.

Rossetto O, Schiavo G, Montecucco C, Poulain B, Deloye F, Lozzi L, and Shone CC (1994). SNARE motif and neurotoxins. *Nature*, **372**, 415–416.

Rossmann MG (1990). The molecular replacement method. *Acta Crystallogr. A*, **46**, 73–82.

Rummel A, Eichner T, Weil T, Karnath T, Gutcaits A, Mahrhold S, Sandhoff K, Proia RL, Acharya KR, Bigalke H, and Binz T (2007). Identification of the protein receptor binding site of botulinum neurotoxins B and G proves the double-receptor concept. *Proc. Natl. Acad. Sci.*, **104**, 359–364.

Rummel A, Häfner K, Mahrhold S, Darashchonak N, Holt M, Jahn R, Beermann S, Karnath T, Bigalke H, and Binz T (2009). Botulinum neurotoxins C, E and F bind gangliosides via a conserved binding site prior to stimulation-dependent uptake with botulinum neurotoxin F utilising the three isoforms of SV2 as second receptor. *J. Neurochem.*, **110**, 1942–1954.

Rummel A, Karnath T, Henke T, Bigalke H, and Binz T (2004a). Synaptotagmins I and II act as nerve cell receptors for botulinum neurotoxin G. *J. Biol. Chem.*, **279**, 30865–30870.

Rummel A, Mahrhold S, Bigalke H, and Binz T (2004b). The HCC-domain of botulinum neurotoxins A and B exhibits a singular ganglioside binding site displaying serotype specific carbohydrate interaction. *Mol. Microbiol.*, **51**, 631–643.

Ruszkowski M and Dauter Z (2016). On methylene-bridged cysteine and lysine residues in proteins. *Protein Sci.*, **25**, 1734–1736.

Sammito M, Meindl K, de Ilarduya IM, Millán C, Artola-Recolons C, Hermoso JA, and Usón I (2014). Structure solution with ARCIMBOLDO using fragments derived from distant homology models. *FEBS J.*, **281**, 4029–4045.

Scheps D, López de la Paz M, Jurk M, Hofmann F, and Frevert J (2017). Design of modified botulinum neurotoxin A1 variants with a shorter persistence of paralysis and duration of action. *Toxicon*, **139**, 101–108.

Schiavo GG, Benfenati F, Poulain B, Rossetto O, de Laureto PP, DasGupta BR, and Montecucco C (1992). Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature*, **359**, 832–835.

Schiavo G, Malizio C, Trimble WS, De Laureto PP, Milan G, Sugiyama H, Johnson EA, and Montecucco C (1994). Botulinum G neurotoxin cleaves VAMP/synaptobrevin at a single Ala-Ala peptide bond. *J. Biol. Chem.*, **269**, 20213–20216.

Schiavo G, Matteoli M, and Montecucco C (2000). Neurotoxins affecting neuroexocytosis. *Physiol. Rev.*, **80**, 717–766.

Schiavo G, Rossetto O, Catsicas S, De Laureto P, DasGupta BR, Benfenati F, and Montecucco C (1993a). Identification of the nerve terminal targets of botulinum neurotoxin serotypes A, D, and E. *J. Biol. Chem.*, **268**, 23784–23787.

Schiavo G, Shone CC, Rossetto O, Alexander FCG, and Montecucco C (1993b). Botulinum neurotoxin serotype F is a zinc endopeptidase specific for VAMP/synaptobrevin. *J. Biol. Chem.*, **268**, 11516–11519.

Schmitt J, Karalewitz A, Benefield DA, Mushrush DJ, Pruitt RN, Spiller BW, Barbieri JT, and Lacy DB (2010). Structural analysis of botulinum neurotoxin type G receptor binding. *Biochemistry*, **49**, 5200–5205.

Schrödinger, LLC (Nov. 2015). "The PyMOL Molecular Graphics System, Version 1.8".

Sebaihia M, Peck MW, Minton NP, Thomson NR, Holden MT, Mitchell WJ, Carter AT, Bentley SD, Mason DR, Crossman L, Paul CJ, Ivens A, Wells-Bennik MH, Davis IJ, Cerdeño-Tárraga AM, Churcher C, Quail MA, Chillingworth T, Feltwell T, Fraser A, Goodhead I, Hance Z,

Jagels K, Larke N, Maddison M, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, White B, Whithead S, and Parkhill J (2007). Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. *Genome Res.*, **17**, 1082–1092.

Shendure J, Balasubramanian S, Church G, Gilbert W, Rogers J, Schloss J, and Waterston R (2017). DNA sequencing at 40: past, present and future. *Nature*, **550**, 345–353.

Shi X, Garcia GE, Nambiar MP, and Gordon RK (2008). Un-nicked BoNT/B activity in human SHSY-5Y neuronal cells. *J. Cell. Biochem.*, **105**, 129–135.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, and Higgins DG (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.

Simons KT, Kooperberg C, Huang E, and Baker D (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.*, **268**, 209–225.

Simpson LL and Rapport MM (1971). Ganglioside inactivation of botulinum toxin. *J. Neurochem.*, **18**, 1341–3.

Smith TJ, Hill KK, Foley BT, Detter JC, Munk AC, Bruce DC, Doggett NA, Smith LA, Marks JD, Xie G, and Brettin TS (2007). Analysis of the neurotoxin complex genes in *Clostridium botulinum* A1-A4 and B1 strains: BoNT/A3, /Ba4 and /B1 clusters are located within plasmids. *PLoS ONE*, **2**.

Sobel J (2005). Botulism. *Clin. Infect. Dis.*, **41**, 1167–1173.

Stefan M, Jasmin S, Consuelo G, Jianlong L, James DM, Andreas R, Thomas B, Mahrhold S, Strotmeier J, Garcia-Rodriguez C, Lou J, Marks JD, Rummel A, and Binz T (2013). Identification of the SV2 protein receptor-binding site of botulinum neurotoxin type E. *Biochem. J.*, **453**, 37–47.

Stenmark P, Dong M, Dupuy J, Chapman ER, and Stevens RC (2010). Crystal structure of the botulinum neurotoxin type G binding domain: insight into cell surface binding. *J. Mol. Biol.*, **397**, 1287–1297.

Stenmark P, Dupuy J, Imamura A, Kiso M, and Stevens RC (2008). Crystal structure of botulinum neurotoxin type A in complex with the cell surface co-receptor GT1b-insight into the toxin-neuron interaction. *PLoS Pathog.*, **4**, e1000129.

Stern D, Weisemann J, Le Blanc A, von Berg L, Mahrhold S, Piesker J, Laue M, Luppa PB, Dorner MB, Dorner BG, and Rummel A (2018). A lipid-binding loop of botulinum neurotoxin serotypes B, DC and G is an essential feature to confer their exquisite potency. *PLoS Pathog.*, **14**, e1007048.

Stiles BG, Pradhan K, Fleming JM, Samy RP, Barth H, and Popoff MR (2014). Clostridium and bacillus binary enterotoxins: bad for the bowels, and eukaryotic being. *Toxins (Basel)*, **6**, 2626–2656.

Stock D, Perisic O, and Löwe J (2005). Robotic nanolitre protein crystallisation at the MRC Laboratory of Molecular Biology. *Prog. Biophys. Mol. Biol.*, **88**, 311–327.

Strotmeier J, Gu S, Jutzi S, Mahrhold S, Zhou J, Pich A, Eichner T, Bigalke H, Rummel A, and Jin R (2011). The biological activity of botulinum neurotoxin type C is dependent upon novel types of ganglioside binding sites. *Mol. Microbiol.*, **81**, 143–156.

Strotmeier J, Lee K, Volker A, Mahrhold S, Zong Y, Zeiser J, Zhou J, Pich A, Bigalke H, and Binz T (2010). Botulinum neurotoxin serotype D attacks neurons via two carbohydrate-binding sites in a ganglioside-dependent manner. *Biochem. J*, **431**, 207–216.

Strotmeier J, Mahrhold S, Krez N, Janzen C, Lou J, Marks JD, Binz T, and Rummel A (2014). Identification of the synaptic vesicle glycoprotein 2 receptor binding site in botulinum neurotoxin A. *FEBS Lett.*, **588**, 1087–1093.

Strotmeier J, Willjes G, Binz T, and Rummel A (2012). Human synaptotagmin-II is not a high affinity receptor for botulinum neurotoxin B and G: Increased therapeutic dosage and immunogenicity. *FEBS Lett.*, **586**, 310–313.

Stuhrmann HB (1980). "Small-angle X-ray scattering of macromolecules in solution". *Synchrotron Radiation Research*. Boston, MA: Springer US, pp. 513–531.

Südhof TC (2002). Synaptotagmins: why so many? *J. Biol. Chem.*, **277**, 7629–7632.

Sun S, Tepp WH, Johnson EA, and Chapman ER (2012). Botulinum neurotoxins B and E translocate at different rates and exhibit divergent responses to GT1b and low pH. *Biochemistry*, **51**, 5655–5662.

Swaminathan S and Eswaramoorthy S (2000). Structural analysis of the catalytic and binding sites of *Clostridium botulinum* neurotoxin B. *Nat. Struct. Biol.*, **7**, 693–699.

Takeda M, Tsukamoto K, Kohda T, Matsui M, Mukamoto M, and Kozaki S (2005). Characterization of the neurotoxin produced by isolates associated with avian botulism. *Avian Diseases*, **49**, 376–381.

Tang J, Maximov A, Shin OH, Dai H, Rizo J, and Sudhof TC (2006). A complexin/synaptotagmin 1 switch controls fast synaptic vesicle exocytosis. *Cell*, **126**, 1175–1187.

Tanizawa Y, Fujisawa T, Mochizuki T, Kaminuma E, Suzuki Y, Nakamura Y, and Tohno M (2014). Draft Genome Sequence of Weissella oryzae SG25T, Isolated from Fermented Rice Grains. *Genome Announc.*, **2**, e00667–14.

Tao L, Peng L, Berntsson RP, Liu SM, Park S, Yu F, Boone C, Palan S, Beard M, Chabrier PE, Stenmark P, Krupp J, and Dong M (2017). Engineered botulinum neurotoxin B with improved efficacy for targeting human receptors. *Nat. Commun.*, **8**, 53.

Tepp WH, Lin G, and Johnson EA (2012). Purification and characterization of a novel subtype a3 botulinum neurotoxin. *Appl. Environ. Microbiol.*, **78**, 3108–3113.

Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Adams PD, Read RJ, Zwart PH, and Hung LW (Apr. 2008). Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallogr. D Biol. Cryst.*, **64**, 515–524.

Tickle IJ, Laskowski RA, and Moss DS (2000). R_{free} and the R_{free} ratio. II. Calculation of the expected values and variances of cross-validation statistics in macromolecular least-squares refinement. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 442–450.

Torii Y, Kiyota N, Sugimoto N, Mori Y, Goto Y, Harakawa T, Nakahira S, Kaji R, Kozaki S, and Ginnaga A (2011). Comparison of effects of botulinum toxin subtype A1 and A2 using twitch tension assay and rat grip strength test. *Toxicon*, **57**, 93–9.

Trewhella J, Duff AP, Durand D, Gabel F, Guss JM, Hendrickson WA, Hura GL, Jacques DA, Kirby NM, Kwan AH, Pérez J, Pollack L, Ryan TM, Sali A, Schneidman-Duhovny D, Schwede T, Svergun DI, Sugiyama M, Tainer JA, Vachette P, Westbrook J, and Whitten AE (2017). 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta. Crystallogr. D. Biol. Crystallogr.*, **73**.

Tsai YC, Kotiya A, Kiris E, Yang M, Bavari S, Tessarollo L, Oyler GA, and Weissman AM (2017). Deubiquitinating enzyme VCIP135 dictates the duration of botulinum neurotoxin type A intoxication. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E5158–E5166.

Tsai YC, Maditz R, Kuo CL, Fishman PS, Shoemaker CB, Oyler GA, and Weissman AM (2010). Targeting botulinum neurotoxin persistence by the ubiquitin-proteasome system. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 16554–9.

Tsukamoto K, Kohda T, Mukamoto M, Takeuchi K, Ihara H, Saito M, and Kozaki S (2005). Binding of *Clostridium botulinum* type C and D neurotoxins to ganglioside and phospholipid: Novel insights into the receptor for clostridial neurotoxins. *J. Biol. Chem.*, **280**, 35164–35171.

Tuukkanen AT and Svergun DI (2014). Weak protein-ligand interactions studied by small-angle X-ray scattering. *FEBS J.*, **281**, 1974–87.

Umeda K, Seto Y, Kohda T, Mukamoto M, and Kozaki S (2009). Genetic characterization of *Clostridium botulinum* associated with type B infant botulism in Japan. *J. Clin. Microbiol.*, **47**, 2720–2728.

Wang J, Meng J, Lawrence GW, Zurawski TH, Sasse A, Bodeker MO, Gilmore MA, Fernández-Salas E, Francis J, Steward LE, Aoki KR, and Dolly JO (2008). Novel chimeras of botulinum neurotoxins A and E unveil contributions from the binding, translocation, and protease domains to their functional characteristics. *J. Biol. Chem.*, **283**, 16993–7002.

Wang J and Wing RA (2014). Diamonds in the rough: A strong case for the inclusion of weak-intensity X-ray diffraction data. *Acta. Crystallogr. D. Biol. Crystallogr.*, **70**, 1491–1497.

Wangroongsarb P, Kohda T, Jittaprasartsin C, Suthivarakom K, Kamthalang T, Umeda K, Sawanpanyalert P, Kozaki S, and Ikuta K (2014). Molecular characterization of *Clostridium botulinum* isolates from foodborne outbreaks in Thailand, 2010. *PLoS ONE*, **9**.

Warkentin M and Thorne RE (2007). A general method for hyperquenching protein crystals. *J. Struct. Funct. Genomics*, **8**, 141–144.

Waterman DG, Winter G, Gildea RJ, Parkhurst JM, Brewster AS, Sauter NK, and Evans G (2016). Diffraction-geometry refinement in the DIALS framework. *Acta Crystallogr. D Struct. Biol.*, **72**, 558–575.

Watts CR, Truong DD, and Nye C (2008). Evidence for the effectiveness of botulinum toxin for spasmodic dysphonia from high-quality research designs. *J. Neural. Transm.*, **115**, 625–630.

Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, and Weiner P (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *JACS*, **106**, 765–784.

Whitemarsh RCM, Tepp WH, Johnson EA, and Pellett S (2014). Persistence of botulinum neurotoxin a subtypes 1-5 in primary rat spinal cord cells. *PLoS ONE*, **9**, 1–6.

Williamson CHD, Vazquez AJ, Hill K, Smith TJ, Nottingham R, Stone NE, Sobek CJ, Cocking JH, Fernández RA, Caballero PA, Leiser OP, Keim P, and Sahl JW (2017). Differentiating botulinum neurotoxin-producing clostridia with a simple, multiplex PCR assay. *Appl. Environ. Microbiol.*, **83**, e00806–17.

Williamson LC, Halpern JL, Montecucco C, Brown JE, and Neale EA (1996). Clostridial neurotoxins and substrate proteolysis in intact neurons: Botulinum neurotoxin C acts on synaptosomal-associated protein of 25 kDa. *J. Biol. Chem.*, **271**, 7694–7699.

Willjes G, Mahrhold S, Strotmeier J, Eichner T, Rummel A, and Binz T (2013). Botulinum neurotoxin G binds synaptotagmin-II in a mode similar to that of serotype B: tyrosine 1186 and lysine 1191 cause its lower affinity. *Biochemistry*, **52**, 3930–8.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, and McCoy A (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.

Winter G, Waterman DG, Parkhurst JM, Brewster AS, Gildea RJ, Gerstel M, Fuentes-Montero L, Vollmar M, Michels-Clark T, Young ID, Sauter NK, and Evans G (2018). DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol*, **74**, 85–97.

Wlodawer A, Minor W, Dauter Z, and Jaskolski M (2013). Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J.*, **280**, 5705–5736.

Yao G, Lam KH, Perry K, Weisemann J, Rummel A, and Jin R (2017). Crystal structure of the receptor-binding domain of botulinum neurotoxin type HA, also known as type FA or H. *Toxins*, **9**.

Yao G, Zhang S, Mahrhold S, Lam K, Stern D, Bagramyan K, Perry K, Kalkum M, Rummel A, Dong M, and Jin R (2016). N-linked glycosylation of SV2 is required for binding and uptake of botulinum neurotoxin A. *Nat. Struct. Mol. Biol.*, **23**, 656–662. eprint: 15334406.

Yeh FL, Dong M, Yao J, Tepp WH, Lin G, Johnson EA, and Chapman ER (2010). SV2 mediates entry of tetanus neurotoxin into central neurons. *PLoS Pathog.*, **6**, e1001207.

Yowler BC, Kensinger RD, and Schengrund CL (2002). Botulinum neurotoxin A activity is dependent upon the presence of specific gangliosides in neuroblastoma cells expressing synaptotagmin I. *J. Biol. Chem.*, **277**, 32815–32819.

Zanetti G, Pirazzini M, Binz T, Shone CC, Fillo S, Lista F, Rossetto O, and Montecucco C (2015). Inhibition of botulinum neurotoxins interchain disulfide bond reduction prevents the peripheral neuroparalysis of botulism. *Biochem. Pharmacol.*, **98**, 522–530.

Zhang S, Berntsson RP, Tepp WH, Tao L, Johnson EA, Stenmark P, and Dong M (2017a). Structural basis for the unique ganglioside and cell membrane recognition mechanism of botulinum neurotoxin DC. *Nat. Commun.*, **8**, 1637.

Zhang S, Lebreton F, Mansfield MJ, Miyashita SI, Zhang J, Schwartzman JA, Tao L, Masuyer G, Martínez-Carranza M, Stenmark P, Gilmore MS, Doxey AC, and Dong M (2018). Identification of a Botulinum Neurotoxin-like Toxin in a Commensal Strain of *Enterococcus faecium*. *Cell Host Microbe*, **23**, 169–176.

Zhang S, Masuyer G, Zhang J, Shen Y, Henriksson L, Miyashita SI, Martínez-Carranza M, Dong M, and Stenmark P (2017b). Identification and characterization of a novel botulinum neurotoxin. *Nat. Comms.*, **8**.

Zhang Y, Buchko GW, Qin L, Robinson H, and Varnum SM (2011). Crystal structure of the receptor binding domain of the botulinum C–D mosaic neurotoxin reveals potential roles of lysines 1118 and 1136 in membrane interactions. *Biochem. Biophys. Res. Comms.*, **404**, 407–412.

Zhang Y, Gardberg AS, Edwards TE, Sankaran B, Robinson H, Varnum SM, and Buchko GW (2013). Structural insights into the functional role of the Hcn sub-domain of the receptor-binding domain of the botulinum neurotoxin mosaic serotype C/D. *Biochimie*, **95**, 1379–85.

Zhang Y, Buchko GW, Qin L, Robinson H, and Varnum SM (2010). Structural analysis of the receptor binding domain of botulinum neurotoxin serotype D. *Biochemical and biophysical research communications*, **401**, 498–503.

Zheng H, Cooper DR, Porebski PJ, Shabalin IG, Handing KB, and Minor W (2017). CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallogr. D Struct. Biol.*, **73**, 223–233.

Zornetta I, Azarnia Tehran D, Arrigoni G, Anniballi F, Bano L, Leka O, Zanotti G, Binz T, and Montecucco C (2016). The first non Clostridial botulinum-like toxin cleaves VAMP within the juxtamembrane domain. *Sci. Rep.*, **6**, 30257.

List of Publications

Davies JR, Hackett GS, Liu SM, and Acharya KR (2018a). High resolution crystal structures of the receptor-binding domain of *Clostridium botulinum* neurotoxin serotypes A and FA. *PeerJ*, **6**, e4552.

Davies JR, Liu SM, and Acharya KR (2018b). Variations in the botulinum neurotoxin binding domain and the potential for novel therapeutics. *Toxins (Basel)*, **10**, E421.

Davies JR, Rees J, Liu SM, and Acharya KR (2018c). High resolution crystal structures of *Clostridium botulinum* neurotoxin A3 and A4 binding domains. *J. Struct. Biol.*, **202**, 113–117.

Masuyer G, Davies JR, Moore K, Chaddock JA, and Acharya KR (2015). Structural analysis of *Clostridium botulinum* neurotoxin type D as a platform for the development of targeted secretion inhibitors. *Sci. Rep.*, **5**, 13397.

Results from Davies *et al.* (2018a) and Davies *et al.* (2018c) are incorporated within chapter 3 of this thesis. Parts of sections 1.1 and 1.2 are duplicated verbatim from the review article published as Davies *et al.* (2018b).

Details from the publication, Masuyer *et al.* (2015), were not included within this thesis. Masuyer *et al.* (2015) reports the crystal structure for a functional fragment of botulinum neurotoxin type D, consisting of the catalytic and translocation domains (LHn), and that of SXN101959. SXN101959 is a functional LHn derivative fused to a GHRH-ligand that has previously showed promising results in inhibiting growth hormone release from pituitary somatotrophs. These structures demonstrate the robustness of the LHn fold and highlight their potential for use in engineering targeted secretion inhibitors. The author of this thesis was responsible for the data processing and refinement of SXN101959.

High resolution crystal structures of the receptor-binding domain of *Clostridium botulinum* neurotoxin serotypes A and FA

Jonathan R. Davies[1], Gavin S. Hackett[2], Sai Man Liu[2] and K. Ravi Acharya[1]

[1] Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom
[2] Ipsen Bioinnovation Limited, Abingdon, United Kingdom

ABSTRACT

The binding specificity of botulinum neurotoxins (BoNTs) is primarily a consequence of their ability to bind to multiple receptors at the same time. BoNTs consist of three distinct domains, a metalloprotease light chain (LC), a translocation domain (H_N) and a receptor-binding domain (H_C). Here we report the crystal structure of H_C/FA, complementing an existing structure through the modelling of a previously unresolved loop which is important for receptor-binding. Our H_C/FA structure also contains a previously unidentified disulphide bond, which we have also observed in one of two crystal forms of H_C/A1. This may have implications for receptor-binding and future recombinant toxin production.

Subjects Biochemistry, Biophysics, Microbiology

Keywords SV2, Crystal structure, Botulinum neurotoxin, Targeted secretion inhibitor, FA hybrid, Receptor binding domain

INTRODUCTION

Botulinum neurotoxin (BoNT) is the active agent that causes the deadly condition botulism. It is expressed as a single polypeptide chain of approximately 150 kDa and is cleaved post-translationally to yield an active di-chain molecule held together by a single disulphide bond. The smaller 50 kDa light chain (LC) possesses a single zinc-endopeptidase domain whereas the larger 100 kDa heavy chain is comprised of two domains—a target cell receptor binding (H_C) domain and a translocation (H_N) domain (*Schiavo et al., 1992*; *Montecucco, 1986*). BoNTs are the most poisonous biological substances known to man and their remarkable toxicity is a result of their highly specific and modular mechanism of action. The toxins target neuronal cell membranes through the formation of a dual receptor binding complex (*Montecucco, 1986*) which allows for internalisation by endocytosis (*Colasante et al., 2013*). The H_N domain then translocates the LC into the cytoplasm where it cleaves a specific SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) protein which consequently inhibits vesicle release. Many immunologically distinct BoNTs have been discovered over the years—serotypes /A through to /G; although, only serotypes /A, /B, /E and /F have been reported to cause botulism in humans (*Coffield et al., 1997*).

Submitted 7 February 2018
Accepted 8 March 2018
Published 21 March 2018

Corresponding author
K. Ravi Acharya, bsskra@bath.ac.uk

Academic editor
Vladimir Uversky

Additional Information and Declarations can be found on page 10

DOI **10.7717/peerj.4552**

Each serotype can be further subdivided into different subtypes based on their amino acid sequence (*Rossetto, Pirazzini & Montecucco, 2014*).

The BoNT H_C domain is responsible for targeting the toxin to a specific cell type and the specific receptors involved have been identified for most serotypes. For example, serotype /A binds to a glycosylated luminal domain of the synaptic vesicle 2 (SV2) protein, preferentially to the C isoform (SV2C) (*Dong et al., 2006*; *Mahrhold et al., 2006*), as well as a ganglioside, namely GT1b (*Rummel et al., 2004*; *Yowler & Schengrund, 2004*; *Stenmark et al., 2008*). X-ray crystallography has revealed protein-backbone hydrogen-bond interactions between β-strands of the BoNT/A1 H_C and the fourth luminal domain of SV2C (SV2C-LD4), as well as electrostatic interactions between charged surfaces (*Benoit et al., 2014*). This is supplemented by additional interactions with the N-linked glycans present on the native SV2 molecule—this network of glycan interactions is key for BoNT potency (*Yao et al., 2016*).

BoNTs have been exploited for therapeutic use in many neurological indications such as dystonia and overactive bladder (*Dressler, 2012*). More recently, BoNTs have been re-engineered to target different cell types and treat specific clinical indications (*Masuyer et al., 2014*; *Masuyer et al., 2015*)—these are known as targeted secretion inhibitors (TSIs). A detailed understanding of the structural aspects of the different BoNT subtypes will prove useful in identifying regions of variability that may help uncover conserved mechanisms of binding, which in turn will guide efforts in developing novel BoNT therapeutics. Therefore, high-resolution, three-dimensional structural analysis of naturally occurring BoNTs, their mosaics and subtypes, would be of great value. Of particular interest are the structural differences arising from sequence differences between subtypes.

Recently, a new bivalent strain of *Clostridium botulinum,* IBCA10-7060, was reported to produce BoNT/B2 and a previously unknown BoNT serotype—"BoNT/H" (*Barash & Arnon, 2014*). This novel toxin is a mosaic molecule and is now more commonly referred to as BoNT/FA (as well as BoNT/HA) due to a LC similar to that of BoNT/F5, a H_N domain similar to that of BoNT/F1, and a H_C domain similar to that of BoNT/A1 (*Maslanka et al., 2015*; *Gonzalez-Escalona et al., 2014*; *Kalb et al., 2015*). The crystal structure of the BoNT/FA binding domain was recently reported (*Yao et al., 2017*). Here, we present a new crystal structure of H_C/FA at 1.95 Å resolution which reveals further structural information that is unresolved in the reported structure. Specifically, our structure reveals a loop previously unmodeled due to lack of density, which is of high importance, and we also observe a disulphide bond which was not present within the previous structure. To this end we have produced two crystal forms of H_C/A1 (determined to 1.45 Å and 1.7 Å respectively) differing by the presence of this disulphide bond.

MATERIALS AND METHODS

Cloning and constructs

The genes encoding the binding domain of BoNT/A1 (H_C/A1) and BoNT/FA (H_C/FA) were provided by Ipsen Bioinnovation Ltd. Each was cloned into the pJ401 expression vector (DNA 2.0, Menlo Park, CA, USA) with an N-terminal 6× histidine tag using

standard molecular biology techniques and confirmed by sequencing (Eurofins Genomics, Germany).

Protein expression

His_6-H_C/A1 and His_6-H_C/FA were expressed in *E. coli* strain BL21(DE3) (Novagen, Madison, WI, USA) using the following protocol. A glycerol stock was used to inoculate 100 mL TB medium containing 50 μg/mL kanamycin, and grown at 37 °C for 16 h. From this, 10 mL of culture was used to inoculate 1 L TB medium containing 50 μg/mL kanamycin and grown at 37 °C to an OD_{600} of 0.6. The temperature was reduced to 16 °C and the cells grown to an OD_{600} of 1.0 at which point 0.5 mM isopropyl-d-1-thiogalactopyranoside (IPTG) was added to induce expression. Cells were grown for an additional 16 h at 16 °C and then harvested by centrifugation at 4,000× g for 30 min.

Protein purification

Expression cell pastes were resuspended in 0.5 M NaCl, 50 mM Tris–HCl pH 7.4, 20 mM imidazole and lysed using a Constant Systems cell disruptor at 20 kPSI. Cell debris were removed by centrifugation at 80,000× g for 30 min and the supernatant was filtered through a 0.22 μm membrane syringe filter. The clarified lysate was loaded onto a 5 mL HisTrap column (GE Healthcare, Little Chalfont, UK), washed, and the target protein eluted with 0.5 M imidazole. His_6-H_C/A1 and His_6-H_C/FA were further purified by SEC using a Superdex 200 16/60 column (GE Healthcare, Little Chalfont, UK) and 0.5 M NaCl, 50 mM Tris-HCl pH 7.4. Purified samples were concentrated to 10 mg/mL using a 10 kDa MWCO centrifugal filter (Millipore, Billerica, MA, USA).

Crystallography

Crystals of His_6-H_C/FA and His_6-H_C/A1 were grown at 16 °C using a 1:1 ratio of protein solution (10 mg/mL) to well solution using the sitting-drop vapour-diffusion method—4 M sodium formate, 0.1 M sodium acetate pH 5.5 for the former, and 0.1 M MIB pH 4.0, 25% w/v PEG 1500 for the latter. Crystals were soaked in cryoprotectant (equal volume of reservoir solution and 50% glycerol) before vitrification in liquid nitrogen. Complete datasets were collected on beamline I03 and I04, respectively, at the Diamond Light Source (Didcot, UK). Diffraction images were processed using DIALS (*Gildea et al., 2014*) and scaled using AIMLESS (*Evans & Murshudov, 2013*) from the CCP4 suite (*Winn et al., 2011*). Data collection statistics are summarised in Table 1. A combination of R_{pim} and $CC_{1/2}$ value were used to determine the resolution cut-off of 1.95 Å and 1.45 Å, respectively. Phase information was determined by molecular replacement using PHASER (*McCoy et al., 2007*) and a previous structure of H_C/A1 (PDB: 2VUA; *Stenmark et al., 2008*) as the initial search model. Multiple rounds of structure refinement were performed by manual correction in COOT (*Emsley et al., 2010*) followed by restrained refinement with REFMAC5 (*Murshudov et al., 2011*). Final validation was performed with MolProbity (*Chen et al., 2010*). Secondary structures were annotated using Stride (*Frishman & Argos, 1995*) and figures were prepared using PyMol (Schrödinger, LLC, New York, NY, USA) and CCP4mg (*McNicholas et al., 2011*).

Table 1 Crystallographic data collection and refinement statistics.

	H_C/FA	H_C/A1 (crystal form 1)	H_C/A1 (crystal form 2)
Space Group	P422	$P2_12_12_1$	$P2_1$
Cell dimensions			
a, *b*, *c* (Å)	118.0, 118.0, 173.8	39.8, 107.3, 107.6	61.4, 53.9, 62.7
α, β, γ (°C)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 106.1, 90.0
Resolution (Å)	24.40–1.95 (1.98–1.95)[a]	24.10–1.45 (1.48–1.45)[a]	60.24–1.80 (1.84–1.80)[a]
R_{merge} (%)	20.1 (168.6)[a]	7.6 (69.2)[a]	14.2 (53.9)[a]
R_{meas} (%)	20.8 (176.0)[a]	9.7 (92.0)[a]	15.0 (58.8)[a]
R_{pim} (%)	5.6 (50.2)[a]	5.8 (60.1)[a]	4.6 (22.5)[a]
CC1/2	0.999 (0.832)[a]	0.997 (0.357)[a]	0.962 (0.954)[a]
$<I/\sigma(I)>$	12.6 (2.3)[a]	6.9 (1.3)[a]	14.4 (3.7)[a]
Completeness (%)	100.0 (100.0)[a]	95.3 (95.1)[a]	98.6 (89.1)[a]
Multiplicity	26.0 (23.7)[a]	3.3 (2.7)[a]	20.4 (11.7)[a]
R_{work} (%)	18.0	17.6	18.8
R_{free} (%)	20.9	22.1	22.3
No. of atoms			
Protein	6,907	3,511	3,263
Water	609	421	322
RMSD bond length (Å)	0.007	0.002	0.005
RMSD bond angle (°)	0.89	0.44	0.73
Wilson B factor ($Å^2$)	24.6	15.9	13.4
Average B factors ($Å^2$)			
Protein	28.2, 30.0	21.2	22.5
Solvent	33.8	34.1	28.0
Ramachandran plot			
Favoured (%)	96.7	96.3	96.4
Allowed (%)	3.4	3.5	3.6
Disallowed (%)	0.0	0.2	0.0
PDB code	5MK8	5MK6	5MK7

Notes.
[a]Values in parentheses are for the highest resolution shell.

RESULTS AND DISCUSSION

Crystal structure of BoNT/FA H_C domain

We identified crystallisation conditions which yielded crystals of H_C/FA in space group P422, with two molecules related by non-crystallographic symmetry in the asymmetric unit. This is different to a recently reported structure, PDB: 5V38 (*Yao et al., 2017*) and reveals an important loop that is involved in receptor binding (average temperature factor (B-factor) 74 $Å^2$). A high-multiplicity dataset was collected containing 360 degrees of data over 3,600 images. No significant radiation damage was observed over the course of data collection and thus all data were used. The $CC_{1/2}$ value for the outer shell was 0.832, indicating there was still very usable data at this resolution (*Evans & Murshudov, 2013*).

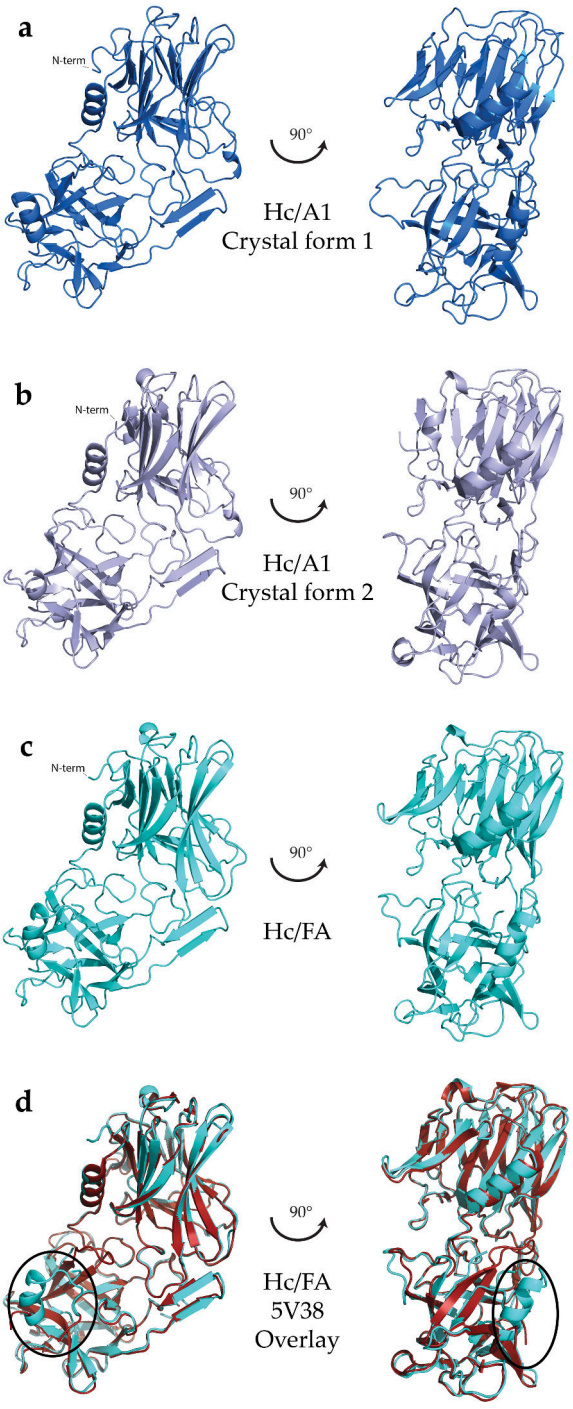


Figure 1 **Crystal structures of HC domains.** (A) HC/A1 domain (crystal form 1, PDB: 5MK6), (B) HC/A1 domain (crystal form 2, PDB: 5MK7), (C) HC/FA domain (PDB: 5MK8), and (D) overlay with a different crystal form of HC/FA (PDB:5V38; *Yao et al., 2017*). The position of loop R1261–R1268 indicated with an ellipse. All structures represented as a ribbon diagram, generated using PyMol (Schrödinger, LLC, New York, NY, USA).

Full-size DOI: 10.7717/peerj.4552/fig-1

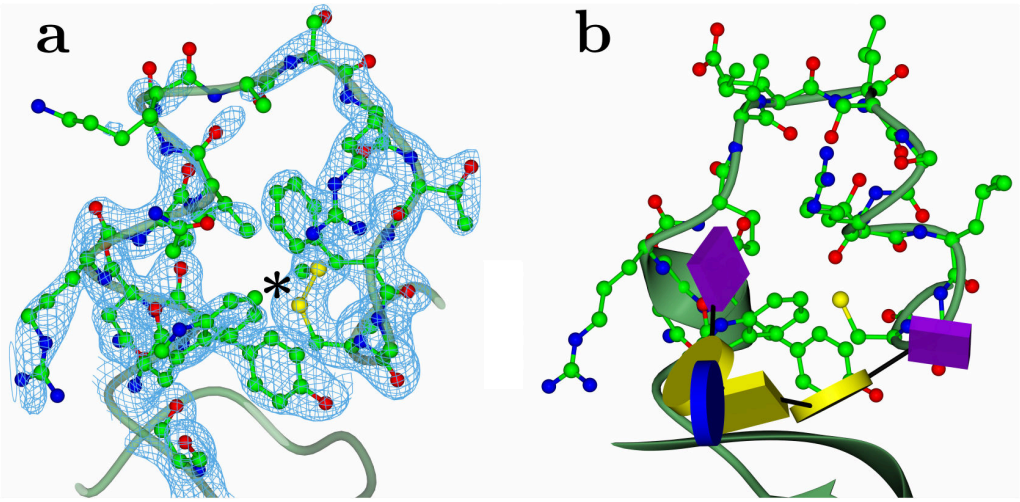


Figure 2 Ganglioside binding site. (A) Electron density from a composite omit map for HC/FA. The location of a disulphide bond is marked with an asterisk. (B) The equivalent loop from HC/A1 (2VU9) with GT1b shown in glycoblock representation. Map produced using Phenix package (*Terwilliger et al., 2008*).
Full-size DOI: 10.7717/peerj.4552/fig-2

The overall structure of H_C/FA is shown in Fig. 1C and the crystallographic statistics are listed in Table 1. As with all reported structures of the BoNT receptor binding domain, present is a characteristic β-jellyroll fold at the N-terminal half and a predominantly β-trefoil fold at the C-terminal half of the protein (Figs. 1A–1D). Both molecules in the asymmetric unit overlay well with a root mean square deviation (RMSD) value of 0.35 Å between all atoms. The B-factors for each chain are low overall (24.74 $Å^2$ and 26.84 $Å^2$ respectively) with a corresponding overall Wilson B-factor of 24.55 $Å^2$. As expected, our structure is highly similar to PDB: 5V38 with an RMSD value for combinations of chains between structures ranges from 0.54 Å to 0.36 Å (Fig. 1D). However, it further reveals the presence of a loop (R1261–R1268) that has been shown in other subtypes to be involved in ganglioside receptor binding (Fig. 2A). Crystal packing has enabled neighbouring chains to interact directly with this loop, provide sufficient stabilisation to produce good electron density. *Yao et al. (2017)* suggested that the lack of density was due to high flexibility, which is consistent with the high B-factors we observed in this region relative to the rest of the protein.

The specific ganglioside receptor for BoNT/FA is not yet known; however, considering that GT1b binds with high affinity to BoNT/A1 (Fig. 2B) and that H_C/FA and H_C/A1 are structurally very similar, we propose that BoNT/FA possesses a similar binding specificity. The overall conformation of this region is also similar to that of BoNT/A1 in complex with a ganglioside receptor (Fig. 2). The detailed conformation of this loop is important for understanding receptor binding and our structure confirms that BoNT/FA could bind to gangliosides in a similar manner to BoNT/A. Proximate to the ganglioside binding region, we observe the presence of a disulphide bond between Cys1227 and Cys1272 (Fig. 3C) which is also not present in the structure 5V38. The equivalent bond has been observed

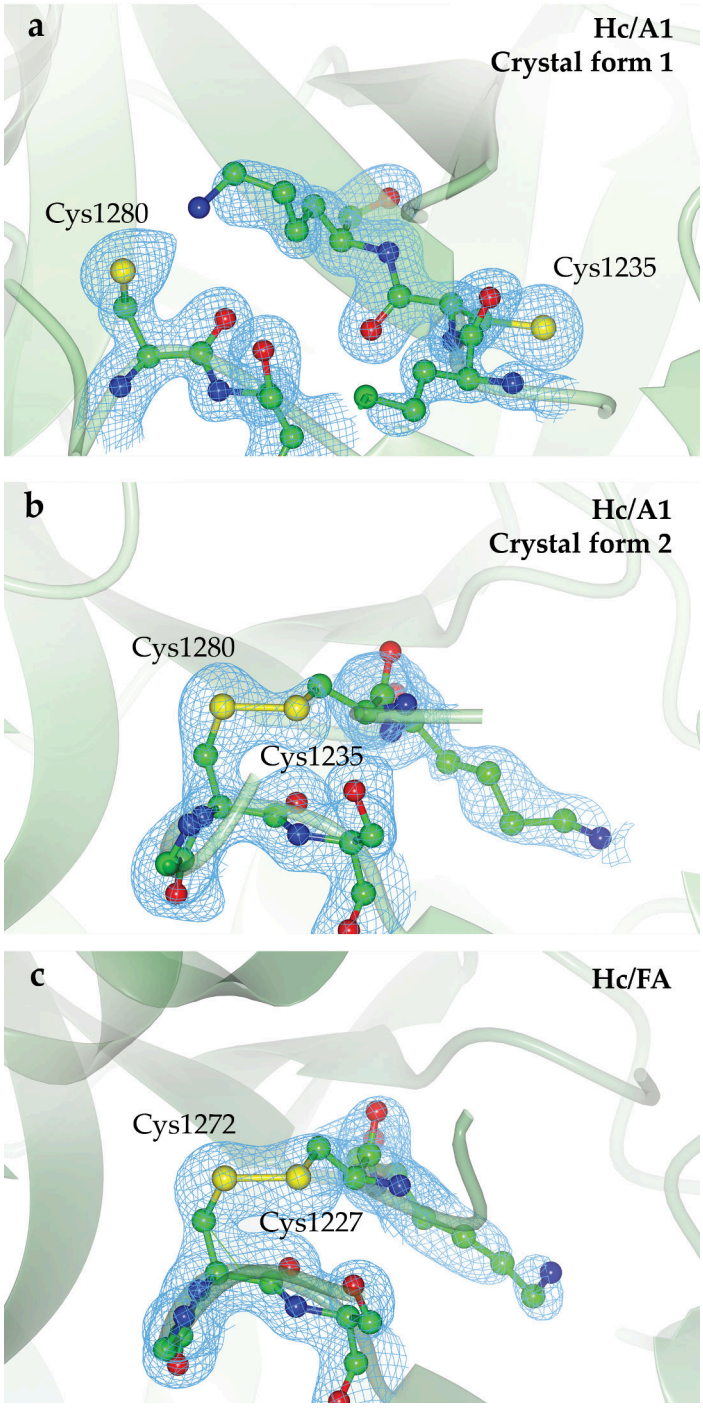


Figure 3 Location of Cys1235 and Cys1280 residues. (A) Cys1235 and Cys1280 are not covalently bound in the HC/A1 crystal form 1, (B) disulphide bond formation in crystal form 2 and (C) a similar disulphide bond is seen in the HC/FA crystal structure between equivalent cysteines. Composite OMIT maps were produced using Phenix (*Terwilliger et al., 2008*). And are shown for each model at 1 σ. Figure generated using CCP4mg (*McNicholas et al., 2011*).

Full-size DOI: 10.7717/peerj.4552/fig-3

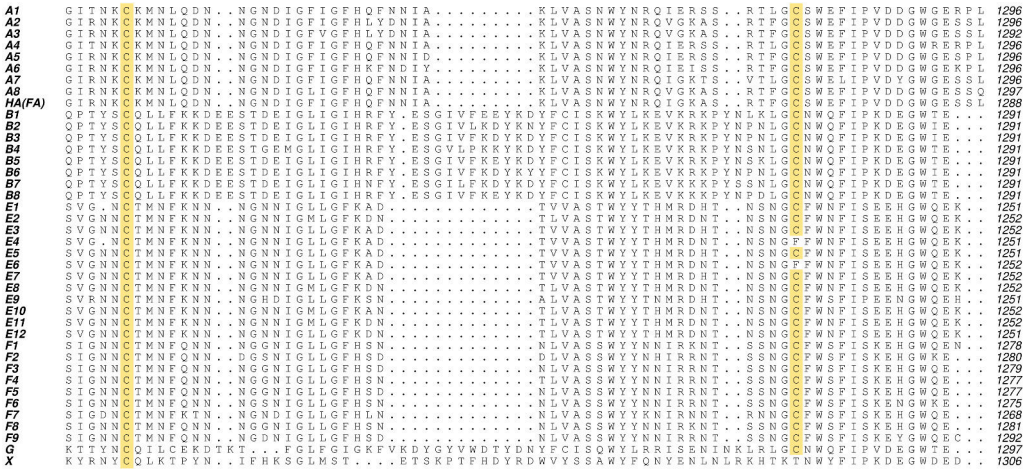


Figure 4 Amino acid sequence alignment of BoNT subtypes emphasising the strong conservation of two cysteine residues near the protein C-terminus. UniprotKB accession numbers—A1: A5HZZ9, A2: Q45894, A3: Q3LRX9, A4: Q3LRX8, A5: C1IPK2, A6: C9WWY7, A7: K4LN57, A8: A0A0A7PDB7, HA(FA): WP_047402807*, B1: P10844, B2: A0A0B4W2B0, B3: A2I2S2, B4: A2I2S4, B5: A0A0E1L271, B6: H3K0G8, B7: H9CNK9, B8: M9VUL2, E1: Q00496, E2: A2I2S6, E3: A0A076L133, E4: P30995, E5: Q9K395, E6: A8Y878, E7: G8I2N7, E8: G8I2N8, E9: WP_017352936*, E10: A0A076JVL9, E11: A0A076K0B0, E12: A0A0A7RCR1, F1: A7GBG3, F2: Q9ZAJ5, F3: D2KHR6, F4: D2KHQ8, F5: D2KHQ9, F6: D2KHS6, F7: D2KHS9, F8: WP_076177537*, F9: A0A1P8YWK9, G: Q60393, X: WP_045538952* (*indicates NCBI accession code where UniprotKB accession is not available).
Full-size DOI: 10.7717/peerj.4552/fig-4

previously in some, but not all crystal structures of BoNT H_C Domains; therefore, it is uncertain what role, if any, it may have towards BoNT function.

Crystal structures of the BoNT/A1 binding domain

We have identified a single crystallisation condition that produced two crystal forms of H_C/A1—one possessed the equivalent disulphide bond whereas the other did not. Using 25% w/v PEG 1500 and 0.1 M MIB pH 4.0, H_C/A1 crystallised into orthorhombic crystals with the space group $P2_12_12_1$ (crystal form 1) that diffracted to a resolution of 1.45 Å. No disulphide bond was observed in this structure (Fig. 1A). Instead, Cys1235 rotated away from the junction due to a backbone rotation, bringing the neighbouring Lys1236 toward Cys1280 (Fig. 3A). This is consistent with previous H_C/A1 structures either in complex with GT1b (PDB: 2VU9; *Stenmark et al., 2008*), SV2-LD (PDB: 4JRA, 5JLV; *Benoit et al., 2014*; *Yao et al., 2016*) or in the apo form (PDB: 2VUA; *Stenmark et al., 2008*). Using the same conditions, monoclinic crystals were obtained six months later with the space group $P2_1$ (crystal form 2) that diffracted to 1.8 Å resolution. Inspection of this structure (Fig. 1B) revealed the presence of a disulphide bond between C1235 and C1280, the equivalent of which was also observed in our H_C/FA crystal structure (Figs. 3A and 3B) and in a full-length BoNT/A1 crystal structure (PDB: 3BTA; *Lacy & Stevens, 1999*). Our findings suggest that the crystallisation condition is not the only determinant as to whether the bond is present or not. The conservation of these cysteine residues across the BoNT sub-serotypes suggests they are very important (Fig. 4); however, their precise

function is unknown. Almost all BoNT subtypes that cause human botulism contain the two residues. They may be required for disulphide bond formation for structural stability or for efficient function of a nearby ganglioside-binding pocket. It must be noted that the expression system does not appear to select for reduced or non-reduced forms of these cysteine residues—each have been observed in different structures of BoNT/A purified from the native *Clostridium botulinum* (*Garcia-Rodriguez et al., 2007*; *Stenmark et al., 2008*; *Fu et al., 2009*; *Gu et al., 2012*; *Przedpelski et al., 2013*; *Benoit et al., 2014*; *Yao et al., 2017*; *Davies et al., 2017*).

CONCLUSION

The high-resolution crystal structures of the binding domains (H_C) from BoNT/FA and BoNT/A1 are reported here. The former complements a recently published structure (*Yao et al., 2017*) and resolves a loop which is highly important for receptor binding. For the latter, two structures were determined from two crystal forms obtained from the same crystallisation condition. These H_C/A1 structures differed from one another by the presence or absence of a disulphide bond. This bond was also observed in our structure of H_C/FA. Considering their location near the ganglioside-binding pocket, and conservation across BoNT subtypes, the redox status of these conserved cysteines may have implications in BoNT stability and manufacture. Botulinum neurotoxins are used therapeutically for many indications and their production is currently from the native host *Clostridium botulinum*. However, considering the safety implications and the advent of engineered BoNT derivatives, such as TSIs, production from a recombinant source would be highly desirable. The importance of these two cysteine residues is being further investigated.

Abbreviations

BoNT/X	botulinum neurotoxin serotype X
BoNT/FA	botulinum neurotoxin hybrid of /F and /A serotypes
H_C	receptor binding domain of BoNT
LC	light chain of BoNT, catalytic domain
H_N	translocation domain of BoNT
SV2C	synaptic vesicle 2 protein, isoform C
TSI	targeted secretion inhibitor

ACKNOWLEDGEMENTS

The authors would like to thank Diamond Light Source, Didcot (Oxon, UK) for beamtime (proposals mx8922 and mx12342) and the staff of beamlines i03 and i04 for assistance with crystal testing and data collection.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

K. Ravi Acharya received a Research Fellowship from Ipsen Bioinnovation Limited and Jonathan R. Davies is supported by a joint post-graduate studentship between University of Bath and Ipsen Bioinnovation Limited. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Ipsen Bioinnovation Limited.
Joint post-graduate studentship between University of Bath and Ipsen Bioinnovation Limited.

Competing Interests

Sai Man Liu is an employee of Ipsen Bioinnovation Limited. Gavin S. Hackett is a former employee of Ipsen Bioinnovation Limited.

Author Contributions

- Jonathan R. Davies conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Gavin S. Hackett analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Sai Man Liu and K. Ravi Acharya conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Accession codes: The atomic coordinates and structure factors (codes 5MK6, 5MK7, 5MK8) have been deposited in the Protein Data Bank (http://www.pdb.org).

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.4552#supplemental-information.

REFERENCES

Barash JR, Arnon SS. 2014. A novel strain of *Clostridium botulinum* that produces type B and type H botulinum toxins. *Journal of Infectious Diseases* **209**:183–191 DOI 10.1093/infdis/jit449.

Benoit RM, Frey D, Hilbert M, Kevenaar JT, Wieser MM, Stirnimann CU, McMillan D, Ceska T, Lebon F, Jaussi R, Steinmetz MO, Schertler GF, Hoogenraad CC, Capitani

G, Kammerer RA. 2014. Structural basis for recognition of synaptic vesicle protein 2C by botulinum neurotoxin A. *Nature* **505**:108–111 DOI 10.1038/nature12732.

Chen VB, Arendall 3rd WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **66**:12–21 DOI 10.1107/S0907444909042073.

Coffield JA, Bakry N, Zhang RD, Carlson J, Gomella LG, Simpson LL. 1997. *In vitro* characterization of botulinum toxin types A, C and D action on human tissues: combined electrophysiologic, pharmacologic and molecular biologic approaches. *Journal of Pharmacology and Experimental Therapeutics* **280**:1489–1498.

Colasante C, Rossetto O, Morbiato L, Pirazzini M, Molgo J, Montecucco C. 2013. Botulinum neurotoxin type A is internalized and translocated from small synaptic vesicles at the neuromuscular junction. *Molecular Neurobiology* **48**:120–127 DOI 10.1007/s12035-013-8423-9.

Davies JR, Rees J, Liu SM, Acharya KR. 2017. High resolution crystal structures of *Clostridium botulinum* neurotoxin A3 and A4 binding domains. *Journal of Structural Biology* Epub ahead of print Dec 26 2017 DOI 10.1016/j.jsb.2017.12.010.

Dong M, Yeh F, Tepp WH, Dean C, Johnson EA, Janz R, Chapman ER. 2006. SV2 is the protein receptor for botulinum neurotoxin A. *Science* **312**:592–596 DOI 10.1126/science.1123654.

Dressler D. 2012. Clinical applications of botulinum toxin. *Current Opinion in Microbiology* **15**:325–336 DOI 10.1016/j.mib.2012.05.012.

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66**:486–501 DOI 10.1107/S0907444910007493.

Evans PR, Murshudov GN. 2013. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* **69**:1204–1214 DOI 10.1107/S0907444913000061.

Frishman D, Argos P. 1995. Knowledge-based protein secondary structure assignment. *Proteins* **23**:566–579 DOI 10.1002/prot.340230412.

Fu Z, Chen C, Barbieri JT, Kim JJ, Baldwin MR. 2009. Glycosylated SV2 and gangliosides as dual receptors for botulinum neurotoxin serotype F. *Biochemistry* **48**:5631–5641 DOI 10.1021/bi9002138.

Garcia-Rodriguez C, Levy R, Arndt JW, Forsyth CM, Razai A, Lou J, Geren I, Stevens RC, Marks JD. 2007. Molecular evolution of antibody cross-reactivity for two subtypes of type A botulinum neurotoxin. *Nature Biotechnology* **25**:107–116 DOI 10.1038/nbt1269.

Gildea RJ, Waterman DG, Parkhurst JM, Axford D, Sutton G, Stuart DI, Sauter NK, Evans G, Winter G. 2014. New methods for indexing multi-lattice diffraction data. *Acta Crystallographica Section D: Biological Crystallography* **70**:2652–2666 DOI 10.1107/S1399004714017039.

Gonzalez-Escalona N, Thirunavukkarasu N, Singh A, Toro M, Brown EW, Zink D, Rummel A, Sharma SK. 2014. Draft genome sequence of bivalent Clostridium

botulinum strain IBCA10-7060, encoding botulinum neurotoxin B and a new FA mosaic type. *Genome Announcements* **2**:e01275–14.

Gu S, Rumpel S, Zhou J, Strotmeier J, Bigalke H, Perry K, Shoemaker CB, Rummel A, Jin R. 2012. Botulinum neurotoxin is shielded by NTNHA in an interlocked complex. *Science* **335**:977–981 DOI 10.1126/science.1214270.

Kalb SR, Baudys J, Raphael BH, Dykes JK, Lúquez C, Maslanka SE, Barr JR. 2015. Functional characterization of botulinum neurotoxin serotype H as a hybrid of known serotypes F and A (BoNT F/A). *Analytical Chemistry* **87**:3911–3917 DOI 10.1021/ac504716v.

Lacy DB, Stevens RC. 1999. Sequence homology and structural analysis of the clostridial neurotoxins. *Journal of Molecular Biology* **291**:1091–1104 DOI 10.1006/jmbi.1999.2945.

Mahrhold S, Rummel A, Bigalke H, Davletov B, Binz T. 2006. The synaptic vesicle protein 2C mediates the uptake of botulinum neurotoxin A into phrenic nerves. *FEBS Letters* **580**:2011–2014 DOI 10.1016/j.febslet.2006.02.074.

Maslanka SE, Luquez C, Dykes JK, Tepp WH, Pier CL, Pellett S, Raphael BH, Kalb SR, Barr JR, Rao A, Johnson EA. 2015. A novel botulinum neurotoxin, previously reported as serotype H, has a hybrid-like structure with regions of similarity to the structures of serotypes A and F and is neutralized with serotype A antitoxin. *Journal of Infectious Diseases* **213**:379–385 DOI 10.1093/infdis/jiv327.

Masuyer G, Chaddock JA, Foster KA, Acharya KR. 2014. Engineered botulinum neurotoxins as new therapeutics. *Annual Review of Pharmacology and Toxicology* **54**:27–51 DOI 10.1146/annurev-pharmtox-011613-135935.

Masuyer G, Davies JR, Moore K, Chaddock JA, Acharya KR. 2015. Structural analysis of *Clostridium botulinum* neurotoxin type D as a platform for the development of targeted secretion inhibitors. *Scientific Reports* **5**:13397 DOI 10.1038/srep13397.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *Journal of Applied Crystallography* **40**:658–674 DOI 10.1107/S0021889807021206.

McNicholas S, Potterton E, Wilson KS, Noble MEM. 2011. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography* **67**:386–394 DOI 10.1107/S0907444911007281.

Montecucco C. 1986. How do tetanus and botulinum toxins bind to neuronal membranes? *TiBS* **11**:314–317.

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* **67**:355–367 DOI 10.1107/S0907444911001314.

Przedpelski A, Tepp WH, Kroken AR, Fu Z, Kim JJ, Johnson EA, Barbieri JT. 2013. Enhancing the protective immune response against botulism. *Infection and Immunity* **81**:2638–2644 DOI 10.1128/IAI.00382-13.

Rossetto O, Pirazzini M, Montecucco C. 2014. Botulinum neurotoxins: genetic, structural and mechanistic insights. *Nature Reviews. Microbiology* **12**:535–549 DOI 10.1038/nrmicro3295.

Rummel A, Mahrhold S, Bigalke H, Binz T. 2004. The HCC-domain of botulinum neurotoxins A and B exhibits a singular ganglioside binding site displaying serotype specific carbohydrate interaction. *Molecular Microbiology* **51**:631–643.

Schiavo G, Rossetto O, Santucci A, DasGupta B, Montecucco C. 1992. Botulinum neurotoxins are zinc proteins. *Journal of Biological Chemistry* **267**:23479–23483.

Stenmark P, Dupuy J, Imamura A, Kiso M, Stevens RC. 2008. Crystal structure of botulinum neurotoxin type A in complex with the cell surface co-receptor GT1b-insight into the toxin-neuron interaction. *PLOS Pathogens* **4**:e1000129 DOI 10.1371/journal.ppat.1000129.

Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Adams PD, Read RJ, Zwart PH, Hung LW. 2008. Iterative-build OMIT maps: map improvement by iterative model building and refinement without model bias. *Acta Crystallographica Section D: Biological Crystallography* **64**:515–524.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**:235–242 DOI 10.1107/S0907444910045749.

Yao G, Lam KH, Perry K, Weisemann J, Rummel A, Jin R. 2017. Crystal structure of the receptor-binding domain of botulinum neurotoxin type HA, also known as type FA or H. *Toxins* **9**:93–105.

Yao G, Zhang S, Mahrhold S, Lam KH, Stern D, Bagramyan K, Perry K, Kalkum M, Rummel A, Dong M, Jin R. 2016. N-linked glycosylation of SV2 is required for binding and uptake of botulinum neurotoxin A. *Nature Structural & Molecular Biology* **23**:656–662 DOI 10.1038/nsmb.3245.

Yowler BC, Schengrund CL. 2004. Glycosphingolipids-sweets for botulinum neurotoxin. *Glycoconjugate Journal* **21**:287–293 DOI 10.1023/B:GLYC.0000046271.64647.fd.

toxins

MDPI

Review

Variations in the Botulinum Neurotoxin Binding Domain and the Potential for Novel Therapeutics

Jonathan R. Davies [1], Sai Man Liu [2] and K. Ravi Acharya [1,*]

[1] Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK; jd769@bath.ac.uk
[2] Ipsen Bioinnovation Limited, Abingdon OX14 4RY, UK; sai.man.liu@ipsen.com
* Correspondence: k.r.acharya@bath.ac.uk; Tel.: +44-1225-386-238

Received: 21 September 2018; Accepted: 18 October 2018; Published: 20 October 2018

Abstract: Botulinum neurotoxins (BoNTs) are categorised into immunologically distinct serotypes BoNT/A to /G). Each serotype can also be further divided into subtypes based on differences in amino acid sequence. BoNTs are ~150 kDa proteins comprised of three major functional domains: an N-terminal zinc metalloprotease light chain (LC), a translocation domain (H_N), and a binding domain (H_C). The H_C is responsible for targeting the BoNT to the neuronal cell membrane, and each serotype has evolved to bind via different mechanisms to different target receptors. Most structural characterisations to date have focussed on the first identified subtype within each serotype (e.g., BoNT/A1). Subtype differences within BoNT serotypes can affect intoxication, displaying different botulism symptoms in vivo, and less emphasis has been placed on investigating these variants. This review outlines the receptors for each BoNT serotype and describes the basis for the highly specific targeting of neuronal cell membranes. Understanding receptor binding is of vital importance, not only for the generation of novel therapeutics but also for understanding how best to protect from intoxication.

Keywords: botulinum neurotoxins; binding domain; ganglioside; SV2; synaptotagmin; neurones

1. Botulinum Neurotoxins

Botulinum neurotoxins (BoNTs) are produced mainly by *Clostridium botulinum*, under anaerobic conditions [1], and are the causative agent of botulism—a rare disease that can lead to paralysis and eventually death. The *C. botulinum* taxon can be divided into four groups (I, II, III, and IV), based on phenotypic differences between the bacteria [2]. *C. botulinum* group I (proteolytic) and group II (non-proteolytic) are mostly responsible for human botulism, whereas *C. botulinum* group III is responsible for botulism in other animal species, and *C. botulinum* group IV does not appear to cause botulism [2,3]. Across these phenotypes, a range of serologically distinct BoNTs have been identified and classified within different serotypes. Until recently, all BoNTs have been categorised into one of seven serotypes ranging from BoNT/A to BoNT/G. The recent identification of novel BoNTs and BoNT-like proteins, which are not neutralisable by existing anti-sera, has lead to classification that does not currently continue from the classical nomenclature (e.g., BoNT/X). Some *C. botulinum* strains have also been identified which express more than one serotype and/or chimeric neurotoxins (e.g., BoNT/CD and BoNT/DC).

Each BoNT is expressed as a single polypeptide chain of ~150 kDa (Figure 1a), after which it is cleaved post-translationally by a protease to yield an active di-chain molecule consisting of a ~50 kDa light chain (LC) and a ~100 kDa heavy chain (HC) linked by a disulphide bond. Some serotypes are cleaved into a di-chain by an endogenous host protease, while others may be cleaved in the target organism [4,5]. For example, BoNT/A purified from *C. botulinum* culture after 8 h is mostly as a single-chain peptide, but when purified from a 96 h culture it is in the nicked di-chain form [6].

The clostridial protease responsible for cleaving BoNT/A has been partially characterised, but not yet identified [7]. The disulphide bond connecting the LC and HC is vital for the mechanism of intoxication [8,9]. The LC possesses a zinc endopeptidase from the M27 family of peptidases, while the HC consists of two domains—a translocation domain (H_N) and a receptor-binding domain (H_C). The H_C domain can be further divided into two distinct folds: a carboxyl-terminal β-trefoil (H_{CC}) with an amino-terminal lectin-like jelly roll (H_{CN}). The crystal structure of BoNT/A1 shows that the domains are arranged in a "butterfly" arrangement where the LC and H_C are the "wings" attached to the central H_N "body" (Figure 1b) [10]. Each domain appears to be structurally and mechanistically distinct from one another, with the exception of a large loop, termed the belt, which wraps around the LC. The structure of BoNT/B follows the same arrangement of BoNT/A, whereas the structure of BoNT/E appears to adopt a "closed wing" compact conformation where the H_C is rotated around the H_N and LC (Figure 1c) [11,12]. The H_C domain is responsible for targeting the protein to the neuronal membrane by binding to receptors present on the cell surface. All classical BoNT serotypes (/A to /G) bind to one or more polysialogangliosides, and most also bind to a protein receptor and together form a dual-receptor complex [13,14].

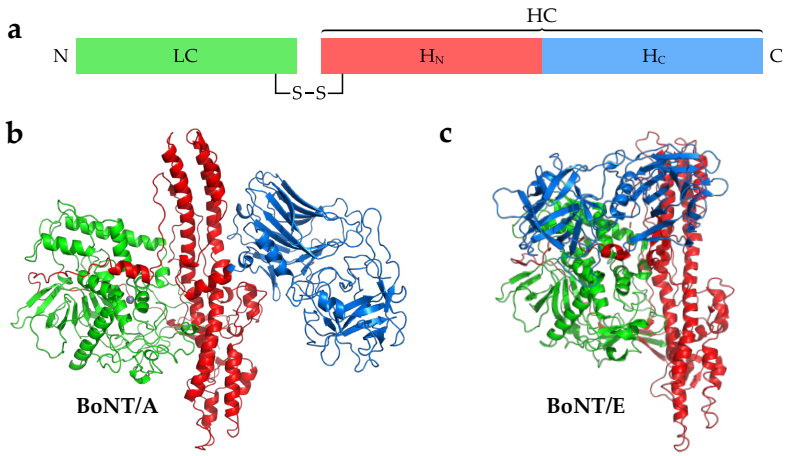


Figure 1. (**a**) Schematic of botulinum neurotoxin (BoNT) domain organisation. The crystal structures of (**b**) BoNT/A [10] and (**c**) BoNT/E [11] show different orientations of the receptor-binding domain (H_C) with respect to the rest of the molecule. H_N: translocation domain; HC: heavy chain; LC: light chain.

Polysialogangliosides consist of a hydrophilic complex polysaccharide with many sialic acid residues, bound to a hydrophobic ceramide tail. Different forms of these gangliosides can be found embedded in the cell membrane with the various sugar moieties displayed on the cell surface. The most common examples found on neuronal membranes include GT1b, GD1a, GD1b, and GM1. Two types of BoNT protein receptors have been identified to date: three isoforms of synaptic vesicle glycoprotein 2 (SV2A-C) and two isoforms of synaptotagmin (SytI-II). Both types are involved with the regulated secretion of neurotransmitter from synaptic vesicles [15,16]. SV2A-C contribute to the modulation of exocytosis, although their exact role is yet to be determined, while SytI and SytII are calcium-sensitive membrane proteins also involved in exocytosis [15,17–19]. Their involvement in synaptic vesicle endocytosis also requires them to be recycled back into the cell through endocytosis, making them excellent targets for BoNTs.

Once the BoNT has bound to its target receptors, it is internalised into a vesicle by endocytosis. The vesicle then matures into an endosome, and proton pumps reduce the internal pH which may cause the BoNT to undergo a conformational change. The exact mechanism of translocation is still not well understood, but it is proposed that the H_N forms a pore through which a partially unfolded LC passes into the neuronal cytosol [20–22]. The LC remains bound to the H_N on the cytosolic side due

to a single disulphide bond, and requires host protein thioredoxin (Trx) and its partner thioredoxin reductase (TrxR) to release the LC (Figure 2). Disulphide cleavage is essential to intoxication, and inhibition of Trx is sufficient to block the LC release [23,24]. The free LC is then able to cleave a soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE), which prevents vesicle–plasma membrane fusion, thus inhibiting exocytosis and release of acetylcholine, causing flaccid paralysis.

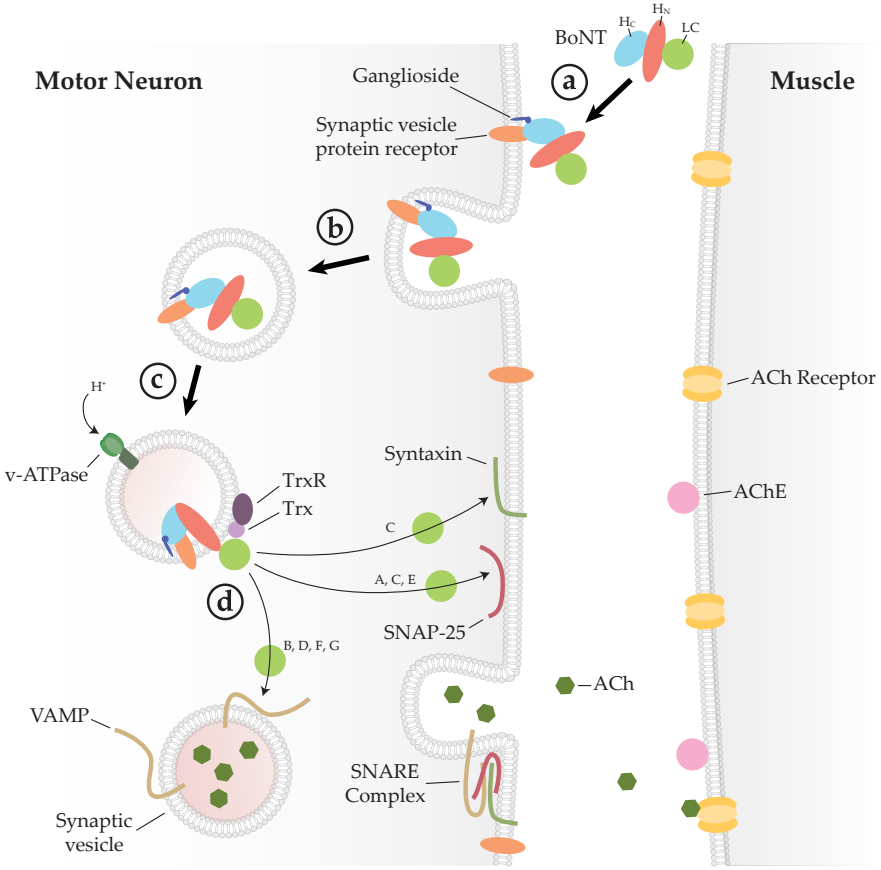


Figure 2. Mechanism of BoNT intoxication. (**a**) BoNT binds the neuronal cell membrane through a dual-receptor complex. (**b**) The BoNT–receptor complex is endocytosed and enclosed within a vesicle. (**c**) Acidification of the endocytic vesicle causes a conformational change, allowing the LC to be translocated through the membrane. (**d**) Thioredoxin (Trx), bound to the vesicle membrane, catalyses the reduction of a disulphide bond that releases the LC into the cytoplasm where it can cleave its soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) substrate. Cleavage of any one of the SNARE proteins of the SNARE complex inhibits membrane fusion and acetylcholine (ACh) release, thus stopping muscle contraction. ACh: acetylcholine; AChE: acetylcholine esterase; Trx: thioredoxin; TrxR: thioredoxin reductase.

2. Receptor-Binding Domain Variation

2.1. BoNT/A

Within the BoNT/A serotype there are currently eight subtypes BoNT/A1 to /A8 which differ by between 3% and 16% at the amino acid level (Table 1). The most thoroughly characterised BoNT subtype is BoNT/A1—this is in part due to its use as a therapeutic for several conditions such as spasticity, dystonias, and glabellar lines [25,26]. The carboxyl-terminal half of the BoNT H_C domain (H_{CC}) contains the peptide motif, H...SxWY...G, which constitutes the core of the ganglioside-binding site (GBS) [27]. In contrast to the dual ganglioside binding sites identified on the related TeNT H_C,

the GBS of BoNT/A1 can only bind one ganglioside at a time [28], but is capable of recognising more than one type, specifically GT1b, GD1a, and to a lesser extent GM1 [29,30]. The exact interactions between ganglioside and the GBS were first determined from the crystal structure of the H_C domain in complex with GT1b [31] (PDB ID: 2VU9). This revealed extensive hydrogen-bonding with four of the seven individual monosaccharides within GT1b. Depletion of gangliosides in neuroblastoma cells completely prevents entry of BoNT/A1 [29]. However, gangliosides alone do not mediate cellular entry—for this, BoNT/A also requires the protein receptor SV2 [32,33]. Of the three SV2 isoforms found in humans, BoNT/A1 has the greatest affinity for SV2C [32]. BoNT/A1 binds specifically to the luminal domain 4 of SV2 (SV2-LD4) via direct backbone–backbone interactions between a β-strand of SV2-LD4 and a β-strand of BoNT H_C [34], and also through interactions with an N559-linked glycan [35]. The significance of the latter is highlighted by the inability of BoNT/A1 to bind to bacterially-expressed (i.e., non-glycosylated) SV2A or SV2B, and a reduced affinity for non-glycosylated SV2C [33,36]. The crystal structure of BoNT/A1–gSV2C-LD4 revealed a large range of interactions between the H_C and the SV2 glycan, which extended away from the backbone–backbone interactions, almost doubling the contact surface area [37].

The other subtypes of BoNT/A are predicted to bind the same receptors as BoNT/A1 due to their high sequence identity between the binding domains (Table 1). The crystal structure of the BoNT/A2 H_C domain in complex with a non-glycosylated SV2C-LD4 showed that the binding mode is conserved, and despite some residue differences, it still binds SV2C [38,39]. F563 of SV2-LD4 forms a π-stacking interaction with R1156 of BoNT/A1, while a glutamic acid residue in BoNT/A2 (E1156) causes F563 to adopt a different conformation and the BoNT instead interacts directly with H564 of SV2C-LD4. Such mutations indicate flexibility with respect to the backbone–backbone interaction of SV2, which may also be tolerated due to extra interactions with the N-linked glycan [37]. BoNT/A2 has also been shown to have a higher affinity for gangliosides than BoNT/A1, although the interactions mediating this difference have not yet been identified [40]. The crystal structures of the BoNT/A3 and /A4 H_C domains suggest a similar mode of interaction to ganglioside compared to BoNT/A1—the GBS of the former shows a potential loss of a hydrogen bond to one of the terminal sialic acids due to a difference in amino acid (phenylalanine rather than a tyrosine), whereas the latter is conformationally conserved [41]. With regard to the SV2 binding site, both structures show slight differences in conformation compared to that of BoNT/A1 due to differences in the primary sequence. Whether this will affect interactions with the SV2 glycan is yet to be determined. Despite high sequence identity between BoNT/A subtypes, significant differences in their intoxication properties have been identified. For example, BoNT/A2 is more potent in neuronal cells than BoNT/A1, possibly due to faster cell entry [42–44], and BoNT/A4 has been reported to be three orders of magnitude less potent than BoNT/A1 [44]. It is difficult to attribute these differences to just interactions between the H_C and receptors, and is instead likely to be as a result of contributions from the H_C, H_N, and LC combined. Uncovering the subtle structural changes resulting from sequence variation which may affect receptor affinity requires further work through structural studies of individual H_C domains and their complexes with receptors.

Table 1. BoNT/A subtype primary sequence identities. Percentage identities are given for full-length and H_C domain sequence alignments of each BoNT/A subtype. For the H_C alignments, sequences aligned to BoNT/A1 residues 870–1296 were used. Uniprot accession codes for BoNT/A1 to /A8 are A5HZZ9 [45–47], Q45894 [48], Q3LRX9 [49], Q3LRX8 [49], C7BEA8 [50], C9WWY7 [51], K4LN57 [52], and A0A0A7PDB7 [53], respectively.

	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	
A1	–	87.29	86.82	91.55	93.90	90.61	91.78	87.79	
A2	90.04	–	98.83	88.47	89.65	90.12	90.35	93.43	
A3	84.66	93.19	–	88.24	88.94	89.65	89.88	92.72	
A4	89.19	88.12	84.29	–	86.85	85.92	85.92	90.14	
A5	97.15	90.50	85.21	87.34	–	93.43	92.72	89.91	
A6	95.68	91.74	86.29	87.72	95.91	–	91.08	87.32	
A7	93.75	89.81	84.90	86.64	94.37	92.98	–	89.67	
A8	93.36	93.44	87.69	88.81	93.60	93.06	91.36	–	
	Full-length identity								**H_C identity**

2.2. BoNT/B

There are currently eight subtypes within the BoNT/B serotype (BoNT/B1 to /B8), and they differ by between 1.5% and 7% at the amino acid level (Table 2) [54]. Although the crystal structure of BoNT/B1 exists in an open conformation similar to that of BoNT/A [11], the BoNT/B serotype targets a different protein receptor on the neuronal cell membrane, namely SytI or SytII [55–58]. Crystal structures of the HC domain from BoNT/B in complex with murine SytII revealed the high specificity of the binding interface where the SytII peptide forms a helix and binds to a hydrophobic groove via six hydrophobic residues [59,60]. Interestingly, BoNT/B displays a much lower affinity toward human SytII than murine SytII due to a single mutation at residue 54-Phe in rodents and Leu in humans [59,61,62]. Considering that SytII is more abundant on human motor neurons than SytI, a significantly larger dose of BoNT/B needs to be administered in order to achieve a similar therapeutic effect to that of BoNT/A. To overcome this issue, the BoNT/B binding domain has been engineered (E1191M,S1199Y) to increase its binding affinity—this molecule showed an 11-fold higher functional efficacy in human cells compared to wild-type BoNT/B1 [63]. BoNT/B is only capable of entering cells once it has bound to both its synaptotagmin receptor and its ganglioside receptor, either GT1b or GD1a [28,55,64]. The crystal structure of the BoNT/B1 binding domain in complex with both SytII and GD1a show strong interactions with the Sia5 moiety [65]. Although there is no direct contact between SytII and GD1a, there is some evidence that each can influence binding to the other, possibly due the spatial arrangement of both binding sites [66]. In addition to the dual receptors, BoNT/B has been reported to interact directly with the cell membrane through an exposed hydrophobic loop ("lipid-binding loop") located between the ganglioside and Syt binding sites on the H_C [67].

Table 2. BoNT/B subtype primary sequence identities. Percentage identities are given for full-length and H_C domain sequence alignments of each BoNT/B subtype. For the H_C alignments, sequences aligned to BoNT/B1 residues 862–1291 were used. UniprotKB accession codes for BoNT/B1 to /B8 are P10844 [68], A2I2R7 [69], A2I2S2 [69], A2I2R6 [69], A2I2R9 [70], A8R089 [71], H9CNK9 [72], I6Z8G9 [54], respectively.

	B1	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	
B1	–	91.86	93.02	89.53	95.35	93.72	90.23	92.33	
B2	95.66	–	96.05	91.16	91.40	96.74	93.02	93.02	
B3	95.97	98.45	–	90.47	92.79	96.28	92.79	92.79	
B4	93.41	94.19	94.03	–	88.60	89.77	90.70	89.07	
B5	96.13	95.20	95.51	92.87	–	93.26	90.23	91.86	
B6	96.20	98.45	98.22	93.49	95.41	–	91.86	92.56	
B7	94.81	95.89	95.74	93.88	94.19	95.20	–	90.70	

Table 2. *Cont.*

	B1	B2	B3	B4	B5	B6	B7	B8
B8	95.51	95.97	95.82	93.34	94.50	95.66	94.58	–
Full-length identity								

2.3. BoNT/C

BoNT/C (specifically known as BoNT/C1) is predominantly associated with botulism in animals rather than humans [2,73]. There are no subtypes of the BoNT/C serotype—only two distinct protein sequences have been identified to-date (UniProtKB: P18640 [74], Q93HT3 [75]) which share 99.9% identity. Perhaps confusingly, there are two other botulinum toxins called "C2 toxin" and "C3" which are not "traditional" neurotoxins, but rather refer to different gene products—a binary AB toxin and an exoenzyme, respectively [76–78]. The mechanisms of cell-binding is of great interest because unlike the majority other BoNTs, no protein receptor for BoNT/C1 has yet been identified [79,80]. Interestingly, while the conserved SxWY ganglioside-binding motif is absent from the H_C domain, BoNT/C1 is still able to bind gangliosides [81]. Indeed, an extended hydrophobic loop termed the "ganglioside binding loop" (GBL) was reported to be essential for neuronal binding, but the specific interactions have yet to be determined [80]. Crystal structures of the BoNT/C1 H_C domain in complex with sialic acid revealed two potential binding sites that are independent of the GBS identified in other BoNTs [82,83].

2.4. BoNT/D

Like BoNT/C, there are no subtypes of BoNT/D, of which there are multiple sequences that share a primary sequence identity of >96%. BoNT/D appears to recognise all three isoforms of SV2 [81,82]. Cells lacking SV2 do not get intoxicated by BoNT/D, but this can be restored by the expression of any of the three SV2 isoforms (SV2A, B, C) [84]. It was further demonstrated that SV2A/B knockout neurones displaying a chimeric form of SV2-LD4 (SV2A, B, or C) alone were unable to mediate BoNT/D entry despite rescuing intoxication for BoNT/A and /E. Mutation of the N537 N-linked glycosylation site also had no effect on BoNT/D entry, despite blocking entry to BoNT/E [84]. This suggests that the SV2 receptor-binding domain in BoNT/D may be distinct from other SV2-interacting BoNTs such as BoNT/A. Gangliosides are also required for BoNT/D cell entry [85], however like BoNT/C, BoNT/D does not contain an SxWY motif in the GBS, although the site is still able to recognise gangliosides [86]. It is also proposed to contain a second binding site termed Sia-1, since mutation of this site results in reduced ganglioside binding [87].

2.5. BoNT/E

There are currently twelve known BoNT/E subtypes (BoNT/E1 to /E12) whose amino acid identities vary by up to 12% (Table 3). The protein receptor for BoNT/E is SV2, although only isoforms SV2B and SV2C are capable of mediating entry [36,88], and in the presence of gangliosides [89]. The SxWy motif is conserved in the BoNT/E GBS, and direct binding of GT1b has been observed [90]. No crystal structures of BoNT/E in complex with receptor or ganglioside have yet been solved. Therefore, the precise molecular basis of their interactions have yet to be determined. The native crystal structure of BoNT/E has been solved, and it reveals a conformation that is significantly different from that of BoNT/A and BoNT/B [12]. In this structure, the HC domain wraps around the toxin, giving the protein more compact shape overall. BoNT/E is capable of entering cells much more quickly than BoNT/A [91], and this domain organisation has been proposed to prime the toxin for translocation, resulting in a faster onset of paralysis [12]. However, investigations using various chimeras of BoNT/A1 and BoNT/E1 showed that the speed of translocation is not affected by the binding domain [92].

Table 3. BoNT/E primary sequence identities. Percentage identities are given for full-length and H_C domain sequence alignments of each BoNT/E subtype. For the H_C alignments, sequences aligned to BoNT/E1 residues 848–1252 were used. Uniprot accession codes for BoNT/E1 to /E12 are K7S1V3 [93], A2I2S6 [93], A2I2S5 [93], C4IHM1 [94], Q9K395 [95], A8Y878 [96], G8I2N7 [93], G8I2N8 [93,97], K7S9Y2 [93], A0A076JVL9 [98], A0A076K0B0 [98], W8FNB6 [99], respectively.

	E1	**E2**	**E3**	**E4**	**E5**	**E6**	**E7**	**E8**	**E9**	**E10**	**E11**	**E12**	
E1	–	97.53	100.0	97.78	92.82	97.28	100.0	92.28	83.17	94.57	93.58	87.87	
E2	99.20	–	97.53	97.28	91.34	96.79	97.53	99.75	84.16	95.80	95.31	88.61	
E3	98.16	97.36	–	97.78	92.82	97.28	100.0	92.28	83.17	94.57	93.58	87.87	
E4	97.28	97.12	95.69	–	90.84	99.51	97.78	97.53	82.67	94.07	93.09	86.88	
E5	96.88	96.40	95.20	95.04	–	90.84	92.82	91.09	83.91	88.86	89.60	90.84	
E6	96.96	96.81	95.93	96.96	94.88	–	97.28	97.04	82.67	93.83	93.09	86.88	
E7	97.92	97.12	97.36	96.25	94.88	96.41	–	97.28	83.17	94.57	93.58	87.87	
E8	96.25	97.04	95.69	96.17	94.16	96.81	98.32	–	83.91	96.05	95.56	88.37	
E9	89.05	89.37	88.73	90.01	89.45	88.25	89.21	89.45	–	83.66	85.15	88.37	
E10	95.37	95.77	94.81	94.97	93.53	95.69	96.88	97.84	89.45	–	97.04	87.38	
E11	93.29	93.85	92.57	92.73	92.01	93.13	93.45	94.41	89.05	95.61	–	87.62	
E12	92.97	93.21	92.65	92.65	93.61	91.21	92.57	92.09	91.45	92.01	91.13	–	
	Full-length identity												**H_C identity**

2.6. BoNT/F

In addition to BoNT/F1, there are eight other BoNT/F subtypes (BoNT/F2 to /F9) which differ by up to 30% sequence identity (Table 4). The exact protein receptor for BoNT/F1 has been reported to be glycosylated SV2 [81,100], but this remains to be established conclusively. For example, one study showed that BoNT/F activity decreased when H_C/A was introduced as a competitor molecule [81], whereas a separate study demonstrated that BoNT/F1 entry in neurones was unaffected by a double SV2A/B knockout in cortical neurones (which have negligible expression of SV2C) [101]. For ganglioside binding, BoNT/F1 requires gangliosides containing an α2,3-linked sialic acid on the terminal galactose (i.e., GT1b or GD1a) [81,100]. The SxWY motif is conserved in BoNT/F, and the crystal structure of the H_C domain from BoNT/F1 in complex with GD1a confirmed the existence of a GBS [102].

Table 4. BoNT/F subtype primary sequence identities. Percentage identities are given for full-length and H_C domain sequence alignments of each BoNT/F subtype. For the H_C alignments, sequences aligned to BoNT/F1 residues 866–1278 were used. Uniprot accession codes for BoNT/F1 to /F9 are A7GBG3 [103], D2KHQ7 [103], D2KHR6 [103], D2KHQ8 [103], D2KHQ9 [103], D2KHS4 [103], D2KHS9 [103], KEJ01913 * [104], A0A1P8YWK9 [105], respectively. * GenBank accession code given where UniProtKB code not available.

	F1	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	
F1	–	82.97	84.63	89.29	83.94	82.51	79.56	98.05	84.71	
F2	83.71	–	96.37	81.51	92.03	93.89	72.02	83.70	90.34	
F3	84.25	97.19	–	82.44	93.46	93.64	73.17	84.39	92.49	
F4	92.33	83.71	84.09	–	82.00	81.03	76.40	88.56	83.45	
F5	70.31	74.37	74.35	69.84	–	90.22	73.48	84.18	92.03	
F6	88.05	90.20	90.04	87.42	74.11	–	72.17	82.27	88.75	
F7	74.43	69.53	69.91	72.77	64.45	70.84	–	79.32	73.48	
F8	96.24	83.71	84.17	93.19	69.84	87.81	73.01	–	84.67	
F9	84.27	89.92	81.63	84.03	73.75	87.37	69.85	84.18	–	
	Full-length identity									**H_C identity**

2.7. BoNT/G

Only two protein sequences of BoNT/G are currently known to exist, and they share 99.9% amino acid identity. The protein receptor for BoNT/G is either SytI or SytII, although interestingly the interface diverges from BoNT/B and it has a lower binding affinity [58,106,107]. Only 5 of 14 residues involved in the BoNT/B–SytII interaction are conserved [57,106]. BoNT/G also displays a low affinity for the human SytII receptor due to a human/chimpanzee-specific mutation [61]. The BoNT/B H_C domain was successfully engineered to improve human SytII binding, and a similar approach would be worth investigating here [63]. BoNT/G possesses the conserved SxWY motif in its GBS, and binds preferentially to GT1b [108]. In addition to the dual-receptor interactions, BoNT/G also contains a "lipid-binding loop" (residues 1252–1256) similar to that of BoNT/B which can directly interact with the cell membrane to further contribute binding affinity [67,106], and deletion of this loop dramatically decreased neurotoxicity [67].

2.8. Mosaic/Chimeric BoNTs

BoNTs composed of domains from different serotypes also exist in nature. The most common of these chimeric toxins are discussed below.

2.8.1. BoNT/CD

BoNT/CD is a mosaic toxin composed of a LC domain and a H_N domain that is most similar to BoNT/C and a H_C domain that is most similar to BoNT/D. Interestingly the binding domain of BoNT/CD binds synaptosomes more tightly than BoNT/D [79]. This may be due to residues K1118 and K1136 (which differ from the equivalent residues in BoNT/D, E1114 and G1132) since mutation of these lysines results in a dramatic loss in synaptosome binding affinity [109]. Protein residues which may also interact with a ganglioside have also been identified through crystallisation with a sialic acid molecule [110].

2.8.2. BoNT/DC

The BoNT/DC chimera possesses an LC domain and a H_N with 96% sequence identity to BoNT/D and a H_C domain similar to that of BoNT/C (74% sequence identity) [111,112]. Botulism caused by BoNT/DC is usually found outside of humans in birds and other mammals, but it is also capable of binding human neuronal cells [62,112]. Despite having a binding domain similar to BoNT/C, BoNT/DC binds to either SytI or SytII. This interaction is mediated by hydrophobic residues, and is distinct from that of BoNT/B [113]. The BoNT/DC protein is particularly interesting, as it appears that it may not require complex gangliosides to enter target neurones [114,115]. However, the crystal structure of BoNT/DC in complex with Sialyl-T suggests that BoNT/DC is capable of recognising a single sialic acid, and thus potentially a range of membrane-bound sugars. The structure also reveals the presence of an extended "lipid-binding loop" that is also observed in BoNT/B and BoNT/G [67,114].

2.8.3. BoNT/HA(FA)

BoNT/FA was recently identified in 2014 from a case of infant botulism [116,117]. At the time it was referred to as BoNT/H (and sometimes still as BoNT/HA) due to its non-neutralisable antigenicity [116], and phylogenetic analysis of the *bont* sequences placed the gene in a lineage distinct from other serotypes [117]. The sequence was finally released to the scientific community after a protracted period of data restriction due to supposed safety concerns [118–120]. It was determined that the molecule was a mosaic toxin composed of an LC similar to that of BoNT/F5, an H_N domain similar to that of BoNT/F1, and an H_C domain similar to that of BoNT/A1 [121]. Direct binding of the BoNT/FA HC domain has been confirmed for glycosylated SV2C-LD4 [122], and crystal structures of this binding domain show some slight differences with respect to BoNT/A1 which would be

consistent with a decreased affinity towards the protein backbone of SV2 [122,123]. Although no ganglioside-bound structure of BoNT/FA has yet been solved, the structure of the GBS appears to maintain the same fold as that observed for BoNT/A1 [123]. SV2 is likely the protein receptor for BoNT/FA, and direct binding has been confirmed for glycosylated SV2C-LD4 [122]. The BoNT/FA sequence contains mutations with respect to BoNT/A1 which result in decreased affinity towards the protein backbone of SV2, as determined by a pull-down assay against non-glycosylated SV2C, while the equivalent residues involved in glycan binding remain unchanged [122]. The effect of these mutations towards different isoforms of SV2 remains to be seen. The ganglioside-binding site is able to maintain the same fold as BoNT/A1, but no ganglioside-bound structures yet exist, so the exact interactions remain to be determined [123]. In recent assays using cultured rat embryonic spinal cord neurones and rat cortical neurones, BoNT/FA was found to be much more potent than BoNT/A1. However, counterintuitively the toxin was much less potent when assayed using an ex vivo mouse phrenic nerve hemidiaphragm (mPNHD). These results, along with the methods used for each assay, point toward a toxin that may have a slow speed of onset despite a highly active LC [124]. Understanding the interactions of BoNT/FA with its receptors is crucial to both determining what causes intoxication differences and for developing novel therapeutics.

2.9. BoNT/X

A strain of *C. botulinum* that was already known to express BoNT/B was recently found to contain the gene for another BoNT molecule that shared low primary sequence identity to other serotypes (<30%)—this was named BoNT/X [125]. It is unknown whether this molecule is capable of causing human botulism, but interestingly, its LC cleaved non-canonical substrates such as VAMP4, VAMP5, and Ykt6 [125]. This suggests that this toxin significantly diverged from other serotypes during its evolution. Despite this, recent structural characterisation of the LC has revealed a core fold common to all BoNTs [126]. Little is known about the BoNT/X H_N and H_C domains, and considering the novel characteristics of LC, attempts are underway to determine the specific receptor(s) that it targets and how it functions in vivo. The BoNT/X H_C does contain an SxWY sequence motif, indicating that it potentially shares similar ganglioside binding characteristics with other BoNTs. Due to its divergence and low sequence similarity to existing BoNTs, structural and functional characterisation could lead to new insights into receptor binding that could be exploited for future therapeutics.

2.10. BoNT-Like Proteins

Considering that BoNTs are the deadliest biological agents that exist, it was surprising to find BoNT-like proteins produced by non-*Clostridium* species. The first was found in 2015 and is referred to as "BoNT/Wo", named after the bacterium that produced it, *Weissella oryzae* SG25 [127,128]. BoNT/Wo cleaves VAMP at a unique location (Trp89–Trp90) [129], but it does not contain any typical BoNT motifs in the receptor-binding domain. This would be consistent with zero reported cases of botulism in humans. Indeed, it has been speculated that BoNT/Wo may instead target SNARE-mediated plant defence systems [128]. More recently, another BoNT-like gene cluster was discovered in the bacterium *Enterococcus faecium*, which is a ubiquitous commensal microorganism commonly found in the gut of mammals. The BoNT-like protein, referred to as BoNT/En or eBoNT/J, possesses many traditional BoNT motifs, including a HExxH zinc-binding motif in the LC and a ganglioside-binding SxWY motif in the HC domain [130,131]. Early studies indicate that rodents do not possesses the receptor(s) for BoNT/En intoxication [130].

3. Conclusions

BoNTs are highly specific and potent exotoxins that are being exploited for therapeutic gain. Our knowledge of the molecular aspects of botulinum neurotoxin, such as mechanism of cell targeting and internalisation, is incomplete and mostly limited to only one or two serotypes (i.e., BoNT/A1 and BoNT/B1). We have yet to fully understand the binding mechanism of others and also how subtle

amino acid differences may result in differences of intoxication (i.e., between subtypes). From what we know so far, X-ray crystallography has suggested that the mechanism of binding is more complex than was initially thought. It is possible that BoNTs may accommodate heterogeneous glycosylation of their protein receptors and target a variety of gangliosides to ensure successful binding to their target cell type. It is not a trivial task to determine how BoNTs bind to their receptors on neuronal cell membranes, especially when trying to replicate the conditions in vivo. With the recent discovery of new BoNTs and BoNT-like molecules in other bacterial species, this raises questions regarding the evolution of the *bont* gene cluster, their ability to be transferred between species, the potential implications for biosafety, and the need for an agreed-upon consistent naming convention to avoid confusion and ambiguity [132,133]. Fast characterisation and the generation of neutralising antibodies against these novel toxins is required. Despite the potential dangers posed, the knowledge may lead to the generation of new and safer therapeutics. In particular, atomic data of the receptor-binding domains from individual subtypes could be used for structural and functional analyses, providing insights for the design of novel BoNTs [134]. In summary, this review highlights the need for further functional and structural characterisation of different BoNT subtypes to improve our understanding of what determines the toxicological differences and how they may be used in therapeutics.

Author Contributions: J.R.D wrote the manuscript. S.M.L. and K.R.A. edited and reviewed the manuscript.

Funding: This research received no external funding.

Acknowledgments: J.R.D. is supported by a joint post-graduate studentship between the University of Bath and Ipsen Bioinnovation Limited.

Conflicts of Interest: S.M.L. is an employee of Ipsen Bioinnovation Limited.

References

1. Hatheway, C.L. Toxigenic clostridia. *Clin. Microbiol. Rev.* **1990**, *3*, 66–98, doi:10.1128/CMR.3.1.66. [CrossRef] [PubMed]
2. Collins, M.D.; East, A.K. Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins. *J. Appl. Microbiol.* **1998**, *84*, 5–17, doi:10.1046/j.1365-2672.1997.00313.x. [CrossRef] [PubMed]
3. Carter, A.T.; Peck, M.W. Genomes, neurotoxins and biology of *Clostridium botulinum* Group I and Group II. *Res. Microbiol.* **2015**, *166*, 303–317, doi:10.1016/j.resmic.2014.10.010. [CrossRef] [PubMed]
4. DasGupta, B.R.; Sugiyama, H. Role of a protease in natural activation of *Clostridium botulinum* neurotoxin. *Infect. Immun.* **1972**, *6*, 587–590.
5. Prabakaran, S.; Tepp, W.; DasGupta, B.R. Botulinum neurotoxin types B and E: purification, limited proteolysis by endoproteinase Glu-C and pepsin, and comparison of their identified cleaved sites relative to the three-dimensional structure of type A neurotoxin. *Toxicon* **2001**, *39*, 1515–1531. [CrossRef]
6. Dekleva, M.L.; DasGupta, B.R. Nicking of single chain *Clostridium botulinum* type A neurotoxin by an endogenous protease. *Biochem. Biophys. Res. Commun.* **1989**, *162*, 767–772, doi:10.1016/0006-291X(89)92376-0. [CrossRef]
7. Dekleva, M.L.; DasGupta, B.R. Purification and characterization of a protease from *Clostridium botulinum* type A that nicks single-chain type A botulinum neurotoxin into the di-chain form. *J. Bacteriol.* **1990**, *172*, 2498–2503, doi:10.1128/jb.172.5.2498-2503.1990. [CrossRef] [PubMed]
8. Fischer, A.; Montal, M. Crucial role of the disulfide bridge between botulinum neurotoxin light and heavy chains in protease translocation across membranes. *J. Biol. Chem.* **2007**, *282*, 29604–29611, doi:10.1074/jbc.M703619200. [CrossRef] [PubMed]
9. Pirazzini, M.; Rossetto, O.; Bolognese, P.; Shone, C.C.; Montecucco, C. Double anchorage to the membrane and intact inter-chain disulfide bond are required for the low pH induced entry of tetanus and botulinum neurotoxins into neurons. *Cell. Microbiol.* **2011**, *13*, 1731–1743, doi:10.1111/j.1462-5822.2011.01654.x. [CrossRef] [PubMed]
10. Lacy, D.B.; Tepp, W.; Cohen, A.C.; DasGupta, B.R.; Stevens, R.C. Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nat. Struct. Biol.* **1998**, *5*, 898–902, doi:10.1038/2338. [CrossRef] [PubMed]

11. Swaminathan, S.; Eswaramoorthy, S. Structural analysis of the catalytic and binding sites of *Clostridium botulinum* neurotoxin B. *Nat. Struct. Biol.* **2000**, *7*, 693–699, doi:10.1038/78005. [CrossRef] [PubMed]
12. Kumaran, D.; Eswaramoorthy, S.; Furey, W.; Navaza, J.; Sax, M.; Swaminathan, S. Domain organization in *Clostridium botulinum* neurotoxin type E is unique: Its implication in faster translocation. *J. Mol. Biol.* **2009**, *386*, 233–245, doi:10.1016/j.jmb.2008.12.027. [CrossRef] [PubMed]
13. Montecucco, C. How do tetanus and botulinum toxins bind to neuronal membranes? *Trends Biochem. Sci.* **1986**, *11*, 314–317, doi:10.1016/0968-0004(86)90282-3. [CrossRef]
14. Brunger, A.T.; Rummel, A. Receptor and substrate interactions of clostridial neurotoxins. *Toxicon* **2009**, *54*, 550–560, doi:10.1016/j.toxicon.2008.12.027. [CrossRef] [PubMed]
15. Morgans, C.W.; Kensel-Hammes, P.; Hurley, J.B.; Burton, K.; Idzerda, R.; McKnight, G.S.; Bajjalieh, S.M. Loss of the synaptic vesicle protein SV2B results in reduced neurotransmission and altered synaptic vesicle protein expression in the retina. *PLoS ONE* **2009**, *4*, e5230, doi:10.1371/journal.pone.0005230. [CrossRef] [PubMed]
16. Chapman, E.R. Synaptotagmin: A Ca^{2+} sensor that triggers exocytosis? *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 498–508, doi:10.1038/nrm855. [CrossRef] [PubMed]
17. Tang, J.; Maximov, A.; Shin, O.; Dai, H.; Rizo, J.; Sudhof, T.C. A complexin/synaptotagmin 1 switch controls fast synaptic vesicle exocytosis. *Cell* **2006**, *126*, 1175–1187, doi:10.1016/j.cell.2006.08.030. [CrossRef] [PubMed]
18. Chen, C.; Arai, I.; Satterfield, R.; Young, S.M.; Jonas, P. Synaptotagmin 2 Is the Fast Ca^{2+} sensor at a central inhibitory synapse. *Cell Rep.* **2017**, *18*, 723–736, doi:10.1016/j.celrep.2016.12.067. [CrossRef] [PubMed]
19. Südhof, T.C. Synaptotagmins: Why so many? *J. Biol. Chem.* **2002**, *277*, 7629–7632, doi:10.1074/jbc.R100052200. [CrossRef] [PubMed]
20. Koriazova, L.K.; Montal, M. Translocation of botulinum neurotoxin light chain protease through the heavy chain channel. *Nat. Struct. Biol.* **2003**, *10*, 13–18, doi:10.1038/nsb879. [CrossRef] [PubMed]
21. Pirazzini, M.; Azarnia, T.D.; Leka, O.; Zanetti, G.; Rossetto, O.; Montecucco, C. On the translocation of botulinum and tetanus neurotoxins across the membrane of acidic intracellular compartments. *Biochim. Biophys. Acta* **2016**, *1858*, 467–474, doi:10.1016/j.bbamem.2015.08.014. [CrossRef] [PubMed]
22. Fischer, A.; Sambashivan, S.; Brunger, A.T.; Montal, M. Beltless translocation domain of botulinum neurotoxin A embodies a minimum ion-conductive channel. *J. Biol. Chem.* **2012**, *287*, 1657–1661, doi:10.1074/jbc.C111.319400. [CrossRef] [PubMed]
23. Pirazzini, M.; Azarnia, T.D.; Zanetti, G.; Megighian, A.; Scorzeto, M.; Fillo, S.; Shone, C.C.; Binz, T.; Rossetto, O.; Lista, F.; et al. Thioredoxin and its reductase are present on synaptic vesicles, and their inhibition prevents the paralysis induced by botulinum neurotoxins. *Cell Rep.* **2014**, *8*, 1870–1878, doi:10.1016/j.celrep.2014.08.017. [CrossRef] [PubMed]
24. Zanetti, G.; Pirazzini, M.; Binz, T.; Shone, C.C.; Fillo, S.; Lista, F.; Rossetto, O.; Montecucco, C. Inhibition of botulinum neurotoxins interchain disulfide bond reduction prevents the peripheral neuroparalysis of botulism. *Biochem. Pharmacol.* **2015**, *98*, 522–530, doi:10.1016/j.bcp.2015.09.023. [CrossRef] [PubMed]
25. Johnson, E.A. Clostridial toxins as therapeutic agents: Benefits of nature's most toxic proteins. *Annu. Rev. Microbiol.* **1999**, *53*, 551–575, doi:10.1146/annurev.micro.53.1.551. [CrossRef] [PubMed]
26. Montecucco, C.; Molgó, J. Botulinal neurotoxins: Revival of an old killer. *Curr. Opin. Pharmacol.* **2005**, *5*, 274–279, doi:10.1016/j.coph.2004.12.006. [CrossRef] [PubMed]
27. Foster, K.A. Molecular Aspects of Botulinum Neurotoxin. In *Current Topics in Neurotoxicity*; Springer: New York, NY, USA, 2014.
28. Rummel, A.; Mahrhold, S.; Bigalke, H.; Binz, T. The HCC-domain of botulinum neurotoxins A and B exhibits a singular ganglioside binding site displaying serotype specific carbohydrate interaction. *Mol. Microbiol.* **2004**, *51*, 631–643, doi:10.1046/j.1365-2958.2003.03872.x. [CrossRef] [PubMed]
29. Yowler, B.C.; Kensinger, R.D.; Schengrund, C.L. Botulinum neurotoxin A activity is dependent upon the presence of specific gangliosides in neuroblastoma cells expressing synaptotagmin I. *J. Biol. Chem.* **2002**, *277*, 32815–32819, doi:10.1074/jbc.M205258200. [CrossRef] [PubMed]
30. Hamark, C.; Berntsson, R.P.; Masuyer, G.; Henriksson, L.M.; Gustafsson, R.; Stenmark, P.; Widmalm, G. Glycans confer specificity to the recognition of ganglioside receptors by botulinum neurotoxin A. *J. Am. Chem. Soc.* **2017**, *139*, 218–230, doi:10.1021/jacs.6b09534. [CrossRef] [PubMed]

31. Stenmark, P.; Dupuy, J.; Imamura, A.; Kiso, M.; Stevens, R.C. Crystal structure of botulinum neurotoxin type A in complex with the cell surface co-receptor GT1b-insight into the toxin-neuron interaction. *PLoS Pathog.* **2008**, *4*, e1000129, doi:10.1371/journal.ppat.1000129. [CrossRef] [PubMed]
32. Dong, M.; Yeh, F.; Tepp, W.H.; Dean, C.; Johnson, E.A.; Janz, R.; Chapman, E.R. SV2 is the protein receptor for botulinum neurotoxin A. *Science* **2006**, *312*, 592–596, doi:10.1126/science.1123654. [CrossRef] [PubMed]
33. Mahrhold, S.; Rummel, A.; Bigalke, H.; Davletov, B.; Binz, T. The synaptic vesicle protein 2C mediates the uptake of botulinum neurotoxin A into phrenic nerves. *FEBS Lett.* **2006**, *580*, 2011–2014, doi:10.1016/j.febslet.2006.02.074. [CrossRef] [PubMed]
34. Benoit, R.M.; Frey, D.; Hilbert, M.; Kevenaar, J.T.; Wieser, M.M.; Stirnimann, C.U.; McMillan, D.; Ceska, T.; Lebon, F.; Jaussi, R.; et al. Structural basis for recognition of synaptic vesicle protein 2C by botulinum neurotoxin A. *Nature* **2014**, *505*, 108–111, doi:10.1038/nature12732. [CrossRef] [PubMed]
35. Mahrhold, S.; Bergström, T.; Stern, D.; Dorner, B.G.; Åstot, C.; Rummel, A. Only the complex N559-glycan in the synaptic vesicle glycoprotein 2C mediates high affinity binding to botulinum neurotoxin serotype A1. *Biochem. J.* **2016**, *473*, 2645–2654, doi:10.1042/BCJ20160439. [CrossRef] [PubMed]
36. Dong, M.; Liu, H.; Tepp, W.H.; Johnson, E.A.; Janz, R.; Chapman, E.R. Glycosylated SV2A and SV2B mediate the entry of botulinum neurotoxin E into neurons. *Mol. Biol. Cell* **2008**, *19*, 5226–5237, doi:10.1091/mbc.E08-07-0765. [CrossRef] [PubMed]
37. Yao, G.; Zhang, S.; Mahrhold, S.; Lam, K.H.; Stern, D.; Bagramyan, K.; Perry, K.; Kalkum, M.; Rummel, A.; Dong, M.; et al. N-linked glycosylation of SV2 is required for binding and uptake of botulinum neurotoxin A. *Nat. Struct. Mol. Biol.* **2016**, *23*, 656–662, doi:10.1038/nsmb.3245. [CrossRef] [PubMed]
38. Benoit, R.M.; Schärer, M.A.; Wieser, M.M.; Li, X.; Frey, D.; Kammerer, R.A. Crystal structure of the BoNT/A2 receptor-binding domain in complex with the luminal domain of its neuronal receptor SV2C. *Sci. Rep.* **2017**, *7*, 43588, doi:10.1038/srep43588. [CrossRef] [PubMed]
39. Gustafsson, R.; Zhang, S.; Masuyer, G.; Dong, M.; Stenmark, P. Crystal structure of botulinum neurotoxin A2 in complex with the human protein receptor SV2C reveals plasticity in receptor binding. *Toxins (Basel)* **2018**, *10*, E153, doi:10.3390/toxins10040153. [CrossRef] [PubMed]
40. Kroken, A.; Blum, F.; Zuverink, M.; Barbieri, J. Entry of botulinum neurotoxin subtypes A1 and A2 into neurons. *Infect. Immun.* **2017**, *85*, e00795-16, doi:10.1128/IAI.00795-16. [CrossRef] [PubMed]
41. Davies, J.R.; Rees, J.; Liu, S.M.; Acharya, K.R. High resolution crystal structures of *Clostridium botulinum* neurotoxin A3 and A4 binding domains. *J. Struct. Biol.* **2018**, *202*, 113–117, doi:10.1016/j.jsb.2017.12.010. [CrossRef] [PubMed]
42. Pier, C.L.; Chen, C.; Tepp, W.H.; Lin, G.; Janda, K.D.; Barbieri, J.T.; Pellett, S.; Johnson, E.A. Botulinum neurotoxin subtype A2 enters neuronal cells faster than subtype A1. *FEBS Lett.* **2011**, *585*, 199–206, doi:10.1016/j.febslet.2010.11.045. [CrossRef] [PubMed]
43. Torii, Y.; Kiyota, N.; Sugimoto, N.; Mori, Y.; Goto, Y.; Harakawa, T.; Nakahira, S.; Kaji, R.; Kozaki, S.; Ginnaga, A. Comparison of effects of botulinum toxin subtype A1 and A2 using twitch tension assay and rat grip strength test. *Toxicon* **2011**, *57*, 93–99, doi:10.1016/j.toxicon.2010.10.009. [CrossRef] [PubMed]
44. Pellett, S.; Tepp, W.H.; Whitemarsh, R.C.; Bradshaw, M.; Johnson, E.A. In vivo onset and duration of action varies for botulinum neurotoxin A subtypes 1–5. *Toxicon* **2015**, *107*, 37–42, doi:10.1016/j.toxicon.2015.06.021. [CrossRef] [PubMed]
45. Betley, M.J.; Somers, E.; DasGupta, B.R. Characterization of botulinum type a neurotoxin gene: Delineation of the N-terminal encoding region. *Biochem. Biophys. Res. Commun.* **1989**, *162*, 1388–1395, doi:10.1016/0006-291X(89)90828-0. [CrossRef]
46. Gimenez, J.A.; DasGupta, B.R. Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, and 18 kD fragments. *J. Prot. Chem.* **1993**, *12*, 351–363, doi:10.1007/BF01028197. [CrossRef]
47. Smith, T.J.; Hill, K.K.; Foley, B.T.; Detter, J.C.; Munk, A.C.; Bruce, D.C.; Doggett, N.A.; Smith, L.A.; Marks, J.D.; Xie, G.; et al. Analysis of the neurotoxin complex genes in *Clostridium botulinum* A1–A4 and B1 strains: BoNT/A3, /Ba4 and /B1 clusters are located within plasmids. *PLoS ONE* **2007**, *2*, e1271, doi:10.1371/journal.pone.0001271. [CrossRef] [PubMed]
48. Dover, N.; Barash, J.R.; Hill, K.K.; Davenport, K.W.; Teshima, H.; Xie, G.; Arnon, S.S. *Clostridium botulinum* strain Af84 contains three neurotoxin gene clusters: Bont/A2, bont/F4 and bont/F5. *PLoS ONE* **2013**, *8*, e61205, doi:10.1371/journal.pone.0061205. [CrossRef]

49. Jacobson, M.J.; Lin, G.; Raphael, B.; Andreadis, J.; Johnson, E.A. Analysis of neurotoxin cluster genes in *Clostridium botulinum* strains producing botulinum neurotoxin serotype A subtypes. *Appl. Environ. Microbiol.* **2008**, *74*, 2778–2786, doi:10.1128/AEM.02828-07. [CrossRef] [PubMed]
50. Dover, N.; Barash, J.R.; Arnon, S.S. Novel *Clostridium botulinum* toxin gene arrangement with subtype A5 and partial subtype B3 botulinum neurotoxin genes. *J. Clin. Microbiol.* **2009**, *47*, 2349–2350, doi:10.1128/JCM.00799-09. [CrossRef] [PubMed]
51. Luquez, C.; Raphael, B.H.; Maslanka, S.E. Neurotoxin gene clusters in *Clostridium botulinum* type Ab strains. *Appl. Environ. Microbiol.* **2009**, *75*, 6094–6101, doi:10.1128/AEM.01009-09. [CrossRef] [PubMed]
52. Mazuet, C.; Ezan, E.; Volland, H.; Popoff, M.R.; Becher, F. Toxin detection in patients' sera by mass spectrometry during two outbreaks of type A botulism in France. *J. Clin. Microbiol.* **2012**, *50*, 4091–4094, doi:10.1128/JCM.02392-12. [CrossRef] [PubMed]
53. Kull, S.; Schulz, K.M.; Weisemann, J.; Kirchner, S.; Schreiber, T.; Bollenbach, A.; Dabrowski, P.W.; Nitsche, A.; Kalb, S.R.; Dorner, M.B.; et al. Isolation and functional characterization of the novel *Clostridium botulinum* neurotoxin A8 subtype. *PLoS ONE* **2015**, *10*, e0116381, doi:10.1371/journal.pone.0116381. [CrossRef] [PubMed]
54. Wangroongsarb, P.; Kohda, T.; Jittaprasartsin, C.; Suthivarakom, K.; Kamthalang, T.; Umeda, K.; Sawanpanyalert, P.; Kozaki, S.; Ikuta, K. Molecular characterization of *Clostridium botulinum* isolates from foodborne outbreaks in Thailand, 2010. *PLoS ONE* **2014**, *9*, e77792, doi:10.1371/journal.pone.0077792. [CrossRef] [PubMed]
55. Nishiki, T.; Tokuyama, Y.; Kamata, Y.; Nemoto, Y.; Yoshida, A.; Sekiguchi, M.; Takahashi, M.; Kozaki, S. Binding of botulinum type B neurotoxin to Chinese hamster ovary cells transfected with rat synaptotagmin II cDNA. *Neurosci. Lett.* **1996**, *208*, 105–108, doi:10.1016/0304-3940(96)12557-X. [CrossRef]
56. Dong, M.; Richards, D.A.; Goodnough, M.C.; Tepp, W.H.; Johnson, E.A.; Chapman, E.R. Synaptotagmins I and II mediate entry of botulinum neurotoxin B into cells. *J. Cell Biol.* **2003**, *162*, 1293–1303, doi:10.1083/jcb.200305098. [CrossRef] [PubMed]
57. Rummel, A.; Eichner, T.; Weil, T.; Karnath, T.; Gutcaits, A.; Mahrhold, S.; Sandhoff, K.; Proia, R.L.; Acharya, K.R.; Bigalke, H.; et al. Identification of the protein receptor binding site of botulinum neurotoxins B and G proves the double-receptor concept. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 359–364, doi:10.1073/pnas.0609713104. [CrossRef] [PubMed]
58. Dong, M.; Tepp, W.H.; Liu, H.; Johnson, E.A.; Chapman, E.R. Mechanism of botulinum neurotoxin B and G entry into hippocampal neurons. *J. Cell Biol.* **2007**, *179*, 1511–1522, doi:10.1083/jcb.200707184. [CrossRef] [PubMed]
59. Chai, Q.; Arndt, J.W.; Dong, M.; Tepp, W.H.; Johnson, E.A.; Chapman, E.R.; Stevens, R.C. Structural basis of cell surface receptor recognition by botulinum neurotoxin B. *Nature* **2006**, *444*, 1096, doi:10.1038/nature05411. [CrossRef] [PubMed]
60. Jin, R.; Rummel, A.; Binz, T.; Brunger, A. Botulinum neurotoxin B recognizes its protein receptor with high affinity and specificity. *Nature* **2006**, *444*, 1092–1095, doi:10.1038/nature05387. [CrossRef] [PubMed]
61. Strotmeier, J.; Willjes, G.; Binz, T.; Rummel, A. Human synaptotagmin-II is not a high affinity receptor for botulinum neurotoxin B and G: increased therapeutic dosage and immunogenicity. *FEBS Lett.* **2012**, *586*, 310–313, doi:10.1016/j.febslet.2011.12.037. [CrossRef] [PubMed]
62. Peng, L.; Berntsson, R.P.; Tepp, W.H.; Pitkin, R.M.; Johnson, E.A.; Stenmark, P.; Dong, M. Botulinum neurotoxin D-C uses synaptotagmin I and II as receptors, and human synaptotagmin II is not an effective receptor for type B, D-C and G toxins. *J. Cell Sci.* **2012**, *125*, 3233–3242, doi:10.1242/jcs.103564. [CrossRef] [PubMed]
63. Tao, L.; Peng, L.; Berntsson, R.P.; Liu, S.M.; Park, S.; Yu, F.; Boone, C.; Palan, S.; Beard, M.; Chabrier, P.; et al. Engineered botulinum neurotoxin B with improved efficacy for targeting human receptors. *Nat. Commun.* **2017**, *8*, 53, doi:10.1038/s41467-017-00064-y. [CrossRef] [PubMed]
64. Kohda, T.; Ihara, H.; Seto, Y.; Tsutsuki, H.; Mukamoto, M.; Kozaki, S. Differential contribution of the residues in C-terminal half of the heavy chain of botulinum neurotoxin type B to its binding to the ganglioside GT1b and the synaptotagmin 2/GT1b complex. *Microb. Pathog.* **2007**, *42*, 72–79, doi:10.1016/j.micpath.2006.10.006. [CrossRef] [PubMed]
65. Berntsson, R.P.; Peng, L.; Dong, M.; Stenmark, P. Structure of dual receptor binding to botulinum neurotoxin B. *Nat. Commun.* **2013**, *4*, 2058, doi:10.1038/ncomms3058. [CrossRef] [PubMed]

66. Atassi, M.Z.; Taruishi, M.; Naqvi, M.; Steward, L.E.; Aoki, K.R. Synaptotagmin II and gangliosides bind independently with botulinum neurotoxin B but each restrains the other. *Protein J.* **2014**, *33*, 278–288, doi:10.1007/s10930-014-9557-y. [CrossRef] [PubMed]
67. Stern, D.; Weisemann, J.; Le Blanc, A.; von Berg, L.; Mahrhold, S.; Piesker, J.; Laue, M.; Luppa, P.B.; Dorner, M.B.; Dorner, B.G.; et al. A lipid-binding loop of botulinum neurotoxin serotypes B, DC and G is an essential feature to confer their exquisite potency. *PLoS Pathog.* **2018**, *14*, e1007048, doi:10.1371/journal.ppat.1007048. [CrossRef] [PubMed]
68. Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P. Molecular cloning of the *Clostridium botulinum* structural gene encoding the type B neurotoxin and determination of its entire nucleotide sequence. *Appl. Environ. Microbiol.* **1992**, *58*, 2345–2354. [PubMed]
69. Hill, K.K.; Smith, T.J.; Helma, C.H.; Ticknor, L.O.; Foley, B.T.; Svensson, R.T.; Brown, J.L.; Johnson, E.A.; Smith, L.A.; Okinaka, R.T.; et al. Genetic diversity among Botulinum Neurotoxin-producing clostridial strains. *J. Bacteriol.* **2007**, *189*, 818–832, doi:10.1128/JB.01180-06. [CrossRef] [PubMed]
70. Kenri, T.; Sekizuka, T.; Yamamoto, A.; Iwaki, M.; Komiya, T.; Hatakeyama, T.; Nakajima, H.; Takahashi, M.; Kuroda, M.; Shibayama, K. Genetic characterization and comparison of *Clostridium botulinum* isolates from botulism cases in Japan between 2006 and 2011. *Appl. Environ. Microbiol.* **2014**, *80*, 6954–6964, doi:10.1128/AEM.02134-14. [CrossRef] [PubMed]
71. Kohda, T.; Nakamura, K.; Hosomi, K.; Torii, Y.; Kozaki, S.; Mukamoto, M. Characterization of the functional activity of botulinum neurotoxin subtype B6. *Microbiol. Immunol.* **2017**, *61*, 482–489, doi:10.1111/1348-0421.12540. [CrossRef] [PubMed]
72. Kalb, S.R.; Baudys, J.; Rees, J.C.; Smith, T.J.; Smith, L.A.; Helma, C.H.; Hill, K.; Kull, S.; Kirchner, S.; Dorner, M.B.; et al. De novo subtype and strain identification of botulinum neurotoxin type B through toxin proteomics. *Anal. Bioanal. Chem.* **2012**, *403*, 215–226, doi:10.1007/s00216-012-5767-3. [CrossRef] [PubMed]
73. Lindström, M.; Nevas, M.; Kurki, J.; Sauna-aho, R.; Latvala-Kiesilä, A.; Pölönen, I.; Korkeala, H. Type C botulism due to toxic feed affecting 52,000 farmed foxes and minks in Finland. *J. Clin. Microbiol.* **2004**, *42*, 4718–4725, doi:10.1128/JCM.42.10.4718-4725.2004. [CrossRef] [PubMed]
74. Hauser, D.; Eklund, M.W.; Kurazono, H.; Binz, T.; Niemann, H.; Gill, D.M.; Boquet, P.; Popoff, M. Nucleotide sequence of *Clostridium botulinum* C1 neurotoxin. *Nucleic Acids Res.* **1990**, *18*, 4924, doi:10.1093/nar/18.16.4924. [CrossRef] [PubMed]
75. Takeda, M.; Tsukamoto, K.; Kohda, T.; Matsui, M.; Mukamoto, M.; Kozaki, S. Characterization of the neurotoxin produced by isolates associated with avian botulism. *Avian Dis.* **2005**, *49*, 376–381, doi:10.1637/7347-022305R1.1. [CrossRef] [PubMed]
76. Stiles, B.G.; Pradhan, K.; Fleming, J.M.; Samy, R.P.; Barth, H.; Popoff, M.R. Clostridium and bacillus binary enterotoxins: Bad for the bowels, and eukaryotic being. *Toxins (Basel)* **2014**, *6*, 2626–2656, doi:10.3390/toxins6092626. [CrossRef]
77. Chellapandi, P.; Prisilla, A. Structure, function and evolution of *Clostridium botulinum* C2 and C3 toxins: Insight to poultry and veterinary vaccines. *Curr. Protein Pept. Sci.* **2017**, *18*, 412–424, doi:10.2174/1389203717666161201203311. [CrossRef] [PubMed]
78. Evans, H.R.; Holloway, D.E.; Sutton, J.M.; Ayriss, J.; Shone, C.C.; Acharya, K.R. C3 exoenzyme from *Clostridium botulinum*: Structure of a tetragonal crystal form and a reassessment of NAD-induced flexure. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1502–1505, doi:10.1107/S0907444904011680. [CrossRef] [PubMed]
79. Tsukamoto, K.; Kohda, T.; Mukamoto, M.; Takeuchi, K.; Ihara, H.; Saito, M.; Kozaki, S. Binding of *Clostridium botulinum* type C and D neurotoxins to ganglioside and phospholipid. Novel insights into the receptor for clostridial neurotoxins. *J. Biol. Chem.* **2005**, *280*, 35164–35171, doi:10.1074/jbc.M507596200. [CrossRef] [PubMed]
80. Kroken, A.R.; Karalewitz, A.P.; Fu, Z.; Baldwin, M.R.; Kim, J.J.; Barbieri, J.T. Unique ganglioside binding by botulinum neurotoxins C and D-SA. *FEBS J.* **2011**, *278*, 4486–4496, doi:10.1111/j.1742-4658.2011.08166.x. [CrossRef] [PubMed]
81. Rummel, A.; Häfner, K.; Mahrhold, S.; Darashchonak, N.; Holt, M.; Jahn, R.; Beermann, S.; Karnath, T.; Bigalke, H.; Binz, T. Botulinum neurotoxins C, E and F bind gangliosides via a conserved binding site prior to stimulation-dependent uptake with botulinum neurotoxin F utilising the three isoforms of SV2 as second receptor. *J. Neurochem.* **2009**, *110*, 1942–1954, doi:10.1111/j.1471-4159.2009.06298.x. [CrossRef] [PubMed]

82. Strotmeier, J.; Gu, S.; Jutzi, S.; Mahrhold, S.; Zhou, J.; Pich, A.; Eichner, T.; Bigalke, H.; Rummel, A.; Jin, R.; et al. The biological activity of botulinum neurotoxin type C is dependent upon novel types of ganglioside binding sites. *Mol. Microbiol.* **2011**, *81*, 143–156, doi:10.1111/j.1365-2958.2011.07682.x. [CrossRef] [PubMed]
83. Karalewitz, A.P.; Fu, Z.; Baldwin, M.R.; Kim, J.J.; Barbieri, J.T. Botulinum neurotoxin serotype C associates with dual ganglioside receptors to facilitate cell entry. *J. Biol. Chem.* **2012**, *287*, 40806–40816, doi:10.1074/jbc.M112.404244. [CrossRef] [PubMed]
84. Peng, L.; Tepp, W.H.; Johnson, E.A.; Dong, M. Botulinum neurotoxin D uses synaptic vesicle protein SV2 and gangliosides as receptors. *PLoS Pathog.* **2011**, *7*, e1002008, doi:10.1371/journal.ppat.1002008. [CrossRef] [PubMed]
85. Strotmeier, J.; Lee, K.; Völker, A.K.; Mahrhold, S.; Zong, Y.; Zeiser, J.; Zhou, J.; Pich, A.; Bigalke, H.; Binz, T.; et al. Botulinum neurotoxin serotype D attacks neurons via two carbohydrate-binding sites in a ganglioside-dependent manner. *Biochem. J.* **2010**, *431*, 207–216, doi:10.1042/BJ20101042. [CrossRef] [PubMed]
86. Zhang, Y.; Buchko, G.W.; Qin, L.; Robinson, H.; Varnum, S.M. Structural analysis of the receptor binding domain of botulinum neurotoxin serotype D. *Biochem. Biophys. Res. Commun.* **2010**, *401*, 498–503, doi:10.1016/j.bbrc.2010.09.063. [CrossRef] [PubMed]
87. Kroken, A.R.; Karalewitz, A.P.; Fu, Z.; Kim, J.J.; Barbieri, J.T. Novel ganglioside-mediated entry of botulinum neurotoxin serotype D into neurons. *J. Biol. Chem.* **2011**, *286*, 26828–26837, doi:10.1074/jbc.M111.254086. [CrossRef] [PubMed]
88. Stefan, M.; Jasmin, S.; Consuelo, G.; Jianlong, L.; James, D.M.; Andreas, R.; Thomas, B.; Mahrhold, S.; Strotmeier, J.; Garcia-Rodriguez, C.; et al. Identification of the SV2 protein receptor-binding site of botulinum neurotoxin type E. *Biochem. J.* **2013**, *453*, 37–47, doi:10.1042/BJ20130391. [CrossRef]
89. Kamata, Y.; Kozaki, S.; Sakaguchi, G.; Iwamori, M.; Nagai, Y. Evidence for direct binding of *Clostridium botulinum* type E derivative toxin and its fragments to gangliosides and free fatty acids. *Biochem. Biophys. Res. Commun.* **1986**, *140*, 1015–1019, doi:10.1016/0006-291X(86)90736-9. [CrossRef]
90. Sun, S.; Tepp, W.H.; Johnson, E.A.; Chapman, E.R. Botulinum neurotoxins B and E translocate at different rates and exhibit divergent responses to GT1b and low pH. *Biochemistry* **2012**, *51*, 5655–5662, doi:10.1021/bi3004928. [CrossRef] [PubMed]
91. Keller, J.E.; Cai, F.; Neale, E.A. Uptake of botulinum neurotoxin into cultured neurons. *Biochemistry* **2004**, *43*, 526–532, doi:10.1021/bi0356698. [CrossRef] [PubMed]
92. Wang, J.; Meng, J.; Lawrence, G.W.; Zurawski, T.H.; Sasse, A.; Bodeker, M.O.; Gilmore, M.A.; Fernández-Salas, E.; Francis, J.; Steward, L.E.; et al. Novel chimeras of botulinum neurotoxins A and E unveil contributions from the binding, translocation, and protease domains to their functional characteristics. *J. Biol. Chem.* **2008**, *283*, 16993–17002, doi:10.1074/jbc.M710442200. [CrossRef] [PubMed]
93. Raphael, B.H.; Lautenschlager, M.; Kalb, S.R.; de Jong, L.I.; Frace, M.; Lúquez, C.; Barr, J.R.; Fernández, R.A.; Maslanka, S.E. Analysis of a unique *Clostridium botulinum* strain from the Southern hemisphere producing a novel type E botulinum neurotoxin subtype. *BMC Microbiol.* **2012**, *12*, 245, doi:10.1186/1471-2180-12-245. [CrossRef] [PubMed]
94. Dykes, J.K.; Lúquez, C.; Raphael, B.H.; McCroskey, L.; Maslanka, S.E. Laboratory investigation of the first case of botulism caused by *Clostridium butyricum* type E toxin in the United States. *J. Clin. Microbiol.* **2015**, *53*, 3363–3365, doi:10.1128/JCM.01351-15. [CrossRef] [PubMed]
95. Wang, X.; Maegawa, T.; Karasawa, T.; Kozaki, S.; Tsukamoto, K.; Gyobu, Y.; Yamakawa, K.; Oguma, K.; Sakaguchi, Y.; Nakamura, S. Genetic analysis of type E botulinum toxin-producing *Clostridium butyricum* strains. *Appl. Environ. Microbiol.* **2000**, *66*, 4992–4997, doi:10.1128/AEM.66.11.4992-4997.2000. [CrossRef] [PubMed]
96. Chen, Y.; Korkeala, H.; Aarnikunnas, J.; Lindström, M. Sequencing the botulinum neurotoxin gene and related genes in *Clostridium botulinum* type E strains reveals orfx3 and a novel type E neurotoxin subtype. *J. Bacteriol.* **2007**, *189*, 8643–8650, doi:10.1128/JB.00784-07. [CrossRef] [PubMed]

97. Macdonald, T.E.; Helma, C.H.; Shou, Y.; Valdez, Y.E.; Ticknor, L.O.; Foley, B.T.; Davis, S.W.; Hannett, G.E.; Kelly-Cirino, C.D.; Barash, J.R.; et al. Analysis of *Clostridium botulinum* serotype E strains by using multilocus sequence typing, amplified fragment length polymorphism, variable-number tandem-repeat analysis, and botulinum neurotoxin gene sequencing. *Appl. Environ. Microbiol.* **2011**, *77*, 8625–8634, doi:10.1128/AEM.05155-11. [CrossRef] [PubMed]
98. Weedmark, K.A.; Lambert, D.L.; Mabon, P.; Hayden, K.L.; Urfano, C.J.; Leclair, D.; Van Domselaar, G.; Austin, J.W.; Corbett, C.R. Two novel toxin variants revealed by whole-genome sequencing of 175 *Clostridium botulinum* type E strains. *Appl. Environ. Microbiol.* **2014**, *80*, 6334–6345, doi:10.1128/AEM.01573-14. [CrossRef] [PubMed]
99. Mazuet, C.; Sautereau, J.; Legeay, C.; Bouchier, C.; Bouvet, P.; Popoff, M.R. An atypical outbreak of food-borne botulism due to *Clostridium botulinum* types B and E from ham. *J. Clin. Microbiol.* **2015**, *53*, 722–726, doi:10.1128/JCM.02942-14. [CrossRef] [PubMed]
100. Fu, Z.; Chen, C.; Barbieri, J.T.; Kim, J.P.; Baldwin, M.R. Glycosylated SV2 and gangliosides as dual receptors for botulinum neurotoxin serotype F. *Biochemistry* **2009**, *48*, 5631–5641, doi:10.1021/bi9002138. [CrossRef] [PubMed]
101. Yeh, F.L.; Dong, M.; Yao, J.; Tepp, W.H.; Lin, G.; Johnson, E.A.; Chapman, E.R. SV2 mediates entry of tetanus neurotoxin into central neurons. *PLoS Pathog.* **2010**, *6*, e1001207, doi:10.1371/journal.ppat.1001207. [CrossRef] [PubMed]
102. Benson, M.A.; Fu, Z.; Kim, J.J.; Baldwin, M.R. Unique ganglioside recognition strategies for clostridial neurotoxins. *J. Biol. Chem.* **2011**, *286*, 34015–34022, doi:10.1074/jbc.M111.272054. [CrossRef] [PubMed]
103. Raphael, B.H.; Choudoir, M.J.; Lúquez, C.; Fernández, R.; Maslanka, S.E. Sequence diversity of genes encoding botulinum neurotoxin type F. *Appl. Environ. Microbiol.* **2010**, *76*, 4805–4812, doi:10.1128/AEM.03109-09. [CrossRef] [PubMed]
104. Giordani, F.; Fillo, S.; Anselmo, A.; Palozzi, A.M.; Fortunato, A.; Gentile, B.; Azarnia, T.D.; Ciammaruconi, A.; Spagnolo, F.; Pittiglio, V.; et al. Genomic characterization of Italian *Clostridium botulinum* group I strains. *Infect. Genet. Evol.* **2015**, *36*, 62–71, doi:10.1016/j.meegid.2015.08.042.. [CrossRef] [PubMed]
105. Sikorra, S.; Skiba, M.; Dorner, M.B.; Weisemann, J.; Weil, M.; Valdezate, S.; Davletov, B.; Rummel, A.; Dorner, B.; Binz, T. Botulinum neurotoxin F subtypes cleaving the VAMP-2 Q-K peptide bond exhibit unique catalytic properties and substrate specificities. *Toxins (Basel)* **2018**, *10*, E311, doi:10.3390/toxins10080311. [CrossRef] [PubMed]
106. Stenmark, P.; Dong, M.; Dupuy, J.; Chapman, E.R.; Stevens, R.C. Crystal structure of the botulinum neurotoxin type G binding domain: Insight into cell surface binding. *J. Mol. Biol.* **2010**, *397*, 1287–1297, doi:10.1016/j.jmb.2010.02.041. [CrossRef] [PubMed]
107. Willjes, G.; Mahrhold, S.; Strotmeier, J.; Eichner, T.; Rummel, A.; Binz, T. Botulinum neurotoxin G binds synaptotagmin-II in a mode similar to that of serotype B: Tyrosine 1186 and lysine 1191 cause its lower affinity. *Biochemistry* **2013**, *52*, 3930–3938, doi:10.1021/bi4003502. [CrossRef] [PubMed]
108. Schmitt, J.; Karalewitz, A.; Benefield, D.A.; Mushrush, D.J.; Pruitt, R.N.; Spiller, B.W.; Barbieri, J.T.; Lacy, D.B. Structural analysis of botulinum neurotoxin type G receptor binding. *Biochemistry* **2010**, *49*, 5200–5205, doi:10.1021/bi100412v. [CrossRef] [PubMed]
109. Zhang, Y.; Buchko, G.W.; Qin, L.; Robinson, H.; Varnum, S.M. Crystal structure of the receptor binding domain of the botulinum C-D mosaic neurotoxin reveals potential roles of lysines 1118 and 1136 in membrane interactions. *Biochem. Biophys. Res. Commun.* **2011**, *404*, 407–412, doi:10.1016/j.bbrc.2010.11.134. [CrossRef] [PubMed]
110. Zhang, Y.; Gardberg, A.S.; Edwards, T.E.; Sankaran, B.; Robinson, H.; Varnum, S.M.; Buchko, G.W. Structural insights into the functional role of the Hcn sub-domain of the receptor-binding domain of the botulinum neurotoxin mosaic serotype C/D. *Biochimie* **2013**, *95*, 1379–1385, doi:10.1016/j.biochi.2013.03.006. [CrossRef] [PubMed]
111. Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K. Mosaic structures of neurotoxins produced from *Clostridium botulinum* types C and D organisms. *Biochim. Biophys. Acta* **1996**, *1307*, 123–126, doi:10.1016/0167-4781(96)00006-1. [CrossRef]
112. Nakamura, K.; Kohda, T.; Umeda, K.; Yamamoto, H.; Mukamoto, M.; Kozaki, S. Characterization of the D/C mosaic neurotoxin produced by *Clostridium botulinum* associated with bovine botulism in Japan. *Vet. Microbiol.* **2010**, *140*, 147–154, doi:10.1016/j.vetmic.2009.07.023. [CrossRef] [PubMed]

113. Berntsson, R.P.; Peng, L.; Svensson, L.M.; Dong, M.; Stenmark, P. Crystal structures of botulinum neurotoxin DC in complex with its protein receptors synaptotagmin I and II. *Structure* **2013**, *21*, 1602–1611, doi:10.1016/j.str.2013.06.026. [CrossRef] [PubMed]
114. Zhang, S.; Berntsson, R.; Tepp, W.; Tao, L.; Johnson, E.; Stenmark, P.; Dong, M. Structural basis for the unique ganglioside and cell membrane recognition mechanism of botulinum neurotoxin DC. *Nat. Commun.* **2017**, *8*, 1637, doi:10.1038/s41467-017-01534-z. [CrossRef] [PubMed]
115. Karalewitz, A.P.; Kroken, A.R.; Fu, Z.; Baldwin, M.R.; Kim, J.J.; Barbieri, J.T. Identification of a unique ganglioside binding loop within botulinum neurotoxins C and D-SA. *Biochemistry* **2010**, *49*, 8117–8126, doi:10.1021/bi100865f. [CrossRef] [PubMed]
116. Barash, J.R.; Arnon, S.S. A novel strain of *Clostridium botulinum* that produces type B and type H botulinum toxins. *J. Infect. Dis.* **2014**, *209*, 183–191, doi:10.1093/infdis/jit449. [CrossRef] [PubMed]
117. Dover, N.; Barash, J.R.; Hill, K.K.; Xie, G.; Arnon, S.S. Molecular characterization of a novel botulinum neurotoxin type H gene. *J. Infect. Dis.* **2014**, *209*, 192–202, doi:10.1093/infdis/jit450. [CrossRef] [PubMed]
118. Hooper, D.C.; Hirsch, M.S. Novel *Clostridium botulinum* Toxin and Dual Use Research of Concern Issues. *J. Infect. Dis.* **2014**, *209*, 167, doi:10.1093/infdis/jit528. [CrossRef] [PubMed]
119. Relman, D.A. "Inconvenient Truths" in the pursuit of scientific knowledge and public health. *J. Infect. Dis.* **2014**, *209*, 170–172, doi:10.1093/infdis/jit529. [CrossRef] [PubMed]
120. Johnson, E.A. Validity of botulinum neurotoxin serotype H. *J. Infect. Dis.* **2014**, *210*, 933–992, doi:10.1093/infdis/jiu211. [CrossRef] [PubMed]
121. Gonzalez-Escalona, N.; Thirunavukkarasu, N.; Singh, A.; Toro, M.; Brown, E.; Zink, D.; Rummel, A.; Sharma, S.K. Draft genome sequence of bivalent *Clostridium botulinum* strain IBCA10-7060, encoding botulinum neurotoxin B and a new FA mosaic type. *Genome Announc.* **2014**, *2*, e01275-14, doi:10.1128/genomeA.01275-14. [CrossRef] [PubMed]
122. Yao, G.; Lam, K.H.; Perry, K.; Weisemann, J.; Rummel, A.; Jin, R. Crystal structure of the receptor-binding domain of botulinum neurotoxin type HA, also known as type FA or H. *Toxins (Basel)* **2017**, *9*, 93, doi:10.3390/toxins9030093. [CrossRef] [PubMed]
123. Davies, J.R.; Hackett, G.S.; Liu, S.M.; Acharya, K.R. High resolution crystal structures of the receptor-binding domain of *Clostridium botulinum* neurotoxin serotypes A and FA. *PeerJ* **2018**, *6*, e4552, doi:10.7717/peerj.4552. [CrossRef] [PubMed]
124. Hackett, G.; Moore, K.; Burgin, D.; Hornby, F.; Gray, B.; Elliott, M.; Mir, I.; Beard, M. Purification and characterization of recombinant botulinum neurotoxin serotype FA, also known as serotype H. *Toxins (Basel)* **2018**, *10*, E195, doi:10.3390/toxins10050195. [CrossRef] [PubMed]
125. Zhang, S.; Masuyer, G.; Zhang, J.; Shen, Y.; Lundin, D.; Henriksson, L.; Miyashita, S.I.; Martínez-Carranza, M.; Dong, M.; Stenmark, P. Identification and characterization of a novel botulinum neurotoxin. *Nat. Commun.* **2017**, *8*, 14130, doi:10.1038/ncomms14130. [CrossRef] [PubMed]
126. Masuyer, G.; Zhang, S.; Barkho, S.; Shen, Y.; Henriksson, L.; Košenina, S.; Dong, M.; Stenmark, P. Structural characterisation of the catalytic domain of botulinum neurotoxin X - high activity and unique substrate specificity. *Sci. Rep.* **2018**, *8*, 4518, doi:10.1038/s41598-018-22842-4. [CrossRef] [PubMed]
127. Tanizawa, Y.; Fujisawa, T.; Mochizuki, T.; Kaminuma, E.; Suzuki, Y.; Nakamura, Y.; Tohno, M. Draft genome sequence of *Weissella oryzae* SG25T, isolated from fermented rice grains. *Genome Announc.* **2014**, *2*, e00667-14, doi:10.1128/genomeA.00667-14. [CrossRef] [PubMed]
128. Mansfield, M.J.; Adams, J.B.; Doxey, A.C. Botulinum neurotoxin homologs in non-Clostridium species. *FEBS Lett.* **2015**, *589*, 342–348, doi:10.1016/j.febslet.2014.12.018. [CrossRef] [PubMed]
129. Zornetta, I.; Azarnia Tehran, D.; Arrigoni, G.; Anniballi, F.; Bano, L.; Leka, O.; Zanotti, G.; Binz, T.; Montecucco, C. The first non *Clostridial botulinum*-like toxin cleaves VAMP within the juxtamembrane domain. *Sci. Rep.* **2016**, *6*, 30257, doi:10.1038/srep30257. [CrossRef] [PubMed]
130. Zhang, S.; Lebreton, F.; Mansfield, M.J.; Miyashita, S.I.; Zhang, J.; Schwartzman, J.A.; Tao, L.; Masuyer, G.; Martínez-Carranza, M.; Stenmark, P.; et al. Identification of a botulinum neurotoxin-like toxin in a commensal strain of *Enterococcus faecium*. *Cell Host Microbe* **2018**, *23*, 169–176.e6, doi:10.1016/j.chom.2017.12.018.. [CrossRef] [PubMed]
131. Brunt, J.; Carter, A.T.; Stringer, S.C.; Peck, M.W. Identification of a novel botulinum neurotoxin gene cluster in Enterococcus. *FEBS Lett.* **2018**, *592*, 310–317, doi:10.1002/1873-3468.12969. [CrossRef] [PubMed]

132. Mansfield, M.; Doxey, A. Genomic insights into the evolution and ecology of botulinum neurotoxins. *Pathog. Dis.* **2018**, *76*, doi:10.1093/femspd/fty040. [CrossRef] [PubMed]
133. Williamson, C.H.D.; Vazquez, A.J.; Hill, K.; Smith, T.J.; Nottingham, R.; Stone, N.E.; Sobek, C.J.; Cocking, J.H.; Fernández, R.; Caballero, P.A.; et al. Differentiating botulinum neurotoxin-producing clostridia with a simple, multiplex PCR assay. *Appl. Environ. Microbiol.* **2017**, *83*, e00806-17, doi:10.1128/AEM.00806-17. [CrossRef] [PubMed]
134. Kammerer, R.A.; Benoit, R.M. Botulinum neurotoxins: New questions arising from structural biology. *Trends Biochem. Sci.* **2014**, *39*, 517–526, doi:10.1016/j.tibs.2014.08.009. [CrossRef] [PubMed]

Journal of Structural Biology 202 (2018) 113–117

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

High resolution crystal structures of *Clostridium botulinum* neurotoxin A3 and A4 binding domains

Jonathan R. Davies[a], Jay Rees[a], Sai Man Liu[b], K. Ravi Acharya[a,*]

[a] Department of Biology and Biochemistry, Claverton Down, University of Bath, Bath BA2 7AY, UK
[b] Ipsen Bioinnovation Limited, 102 Park Lane, Milton Park, Abingdon OX14 4RY, UK

ARTICLE INFO

Keywords:
Clostridium botulinum
Botulinum neurotoxin
Subtypes
Binding domain structure
Crystallography

ABSTRACT

Clostridium botulinum neurotoxins (BoNTs) cause the life-threatening condition, botulism. However, while they have the potential to cause serious harm, they are increasingly being utilised for therapeutic applications. BoNTs comprise of seven distinct serotypes termed BoNT/A through BoNT/G, with the most widely characterised being sub-serotype BoNT/A1. Each BoNT consists of three structurally distinct domains, a binding domain (H_C), a translocation domain (H_N), and a proteolytic domain (LC). The H_C domain is responsible for the highly specific targeting of the neurotoxin to neuronal cell membranes. Here, we present two high-resolution structures of the binding domain of subtype BoNT/A3 (H_C/A3) and BoNT/A4 (H_C/A4) at 1.6 Å and 1.34 Å resolution, respectively. The structures of both proteins share a high degree of similarity to other known BoNT H_C domains whilst containing some subtle differences, and are of benefit to research into therapeutic neurotoxins with novel characteristics.

1. Introduction

Botulinum neurotoxins (BoNTs) comprise a family of proteins produced predominantly by *Clostridium botulinum*. There are seven distinct serotypes termed BoNT/A through BoNT/G that cause the deadly condition botulism in vertebrates (Schiavo et al., 2000, Johnson and Montecucco, 2008, Rossetto et al., 2014). Of these serotypes, /A, /B, /E and /F are most commonly associated with human botulism (Coffield et al., 1997). BoNTs are composed of three major domains that are organised into two chains: a 50 kDa light chain (LC) and a 100 kDa heavy chain (HC), that are linked by a disulphide bond. The LC comprises a single zinc-endopeptidase domain while the HC comprises the other two domains – a translocation domain (H_N) responsible for transporting the LC across a plasma membrane, and a binding domain (H_C) responsible for forming a receptor complex with the target neuronal cell. All BoNTs bind polysialogangliosides found on neuronal cell membranes. In addition, BoNT/A, /D, /E and /F also bind the luminal domain of the transmembrane synaptic vesicle 2 (SV2) protein, while BoNT/B, /DC (a mosaic toxin) and /G instead bind to the luminal domain of the protein synaptotagmin I and II (Syt).

Each BoNT serotype can be further categorised into subtypes based on amino acid sequence identity. For BoNT/A there are currently eight known subtypes (BoNT/A1-8) which share between 84% and 97% sequence identity amongst them. The most widely used therapeutic BoNT is the subtype A1. Some subtypes have been reported to exhibit different properties to BoNT/A1 such as onset of symptoms and duration of action (Tepp et al., 2012; Whitemarsh et al., 2014; Henkel et al., 2009). For example, BoNT/A3 has been shown to be less effectively neutralised by anti-BoNT/A1 antibodies whilst also inducing symptoms of intoxication in mice that are distinct from BoNT/A1 (Mazuet et al., 2010; Tepp et al., 2012). It has been suggested that these differences are due to structural differences within the binding domain of BoNT/A3 (H_C/A3) (Tepp et al., 2012). Here we provide structural insight into the subtype differences of the H_C domains from BoNT/A3 and also BoNT/A4, and provide a plausible explanation on how structural differences may play a role in receptor binding.

2. Materials and methods

2.1. Protein expression and purification

The binding domains of BoNT/A3 (residues 866-1292) and BoNT/A4 (residues 870-1296) were cloned into the pJ401 vector (Atum) from their respective full-length sequences (Genbank: ABA29017.1 and ABA29018.1) with an N-terminal 6xHis tag. Constructs were transformed into *E. coli* BL21 cells, which were grown in 0.5 L LB at 37 °C, and induced at an OD_{600} of 0.6 with 1 mM IPTG at 16 °C for 16 h. For purification, cells expressing H_C/A3 were lysed in 50 mM Tris pH 7.4,

* Corresponding author.
E-mail address: bsskra@bath.ac.uk (K.R. Acharya).

https://doi.org/10.1016/j.jsb.2017.12.010
Received 21 November 2017; Received in revised form 20 December 2017; Accepted 23 December 2017
Available online 26 December 2017

0.5 M NaCl, whereas those expressing H_C/A4 were lysed in 50 mM Tris pH 7.4, 0.2 M NaCl, 10 mM Trehalose. Each protein was captured on a HisTrap column (GE Healthcare) and further purified by size-exclusion chromatography on a Superdex 200 column (GE Healthcare).

2.2. Crystallisation

Protein crystallisation conditions were screened using the sitting-drop vapour diffusion method at 16 °C. H_C/A3 (6 mg/mL) and H_C/A4 (4 mg/mL) were dispensed using an Art Robbins Phoenix crystal screening nano dispenser into 96-well 3-drop Intelliwell plates (Molecular Dimensions, England). Multiple screening kits from Molecular Dimensions were used. H_C/A3 crystals were obtained after one week with 0.1 M MIB pH 4.0 25% w/v PEG 1500 from the PACT premier screen (condition B1). Individual crystals were mounted in a loop and dipped in a solution containing reservoir and 50% glycerol in a 1:1 ratio for cryoprotection before vitrification and storage in liquid nitrogen. H_C/A4 crystals were also obtained after one week, but using the Morpheus II screen (Gorrec, 2015) (Molecular Dimensions, England). The best crystal of approximately 60 μm^3 was produced in 10 mM Spermine tetrahydrochloride, 10 mM Spermidine trihydrochloride, 10 mM 1,4-Diaminobutane dihydrochloride, 10 mM D-L-Ornithine monohydrochloride, 0.1 M MOPSO/bis-tris pH 6.5, 15%(w/v) PEG 3 K, 20%(v/v), 1,2,4-butanetriol, 1%(w/v) NDSB 256 (condition H1). This was mounted in a loop and then flash-cooled for storage in liquid nitrogen.

2.3. X-ray data collection and structure determination

Diffraction data for H_C/A3 and H_C/A4 (7200 images total) were collected using 0.1° oscillations and exposures of 0.02 s at beamline IO3 (Diamond Light Source, Didcot, UK) and processed using DIALS (Waterman et al., 2016). The structures of both proteins were solved by molecular replacement with PHASER (McCoy et al., 2007) using the binding domain of BoNT/A1 (PDB 2VUA; Stenmark et al., 2008) as a search model. Both models were manually fitted using COOT (Emsley et al., 2010) and refined with REFMAC (Murshudov et al., 1997) in the CCP4 suite of programs (Winn et al., 2011). The structures were validated with MolProbity (Chen et al., 2010). Crystallographic data processing and refinement statistics are given in Table 1. Figures were prepared using CCP4MG (McNicholas et al., 2011).

Table 1
X-ray data collection and refinement statistics. Inner and outer shell statistics are given in square and curved brackets respectively.

	H_C/A3	H_C/A4
Diamond Beamline	I03, DLS	I03, DLS
Wavelength (Å)	0.9763	0.9763
Crystallographic Statistics		
Space Group	$P2_12_12_1$	$P2_13$
Unit-cell parameters:		
a,b,c (Å)	39.8, 96.5, 111.0	116.0, 116.0, 116.0
$\alpha = \beta = \gamma$ (°)	90	90
Resolution (Å)	[96.47–8.76] 96.47–1.60 (1.63–1.60)	[81.95–7.33] 81.95–1.34 (1.36–1.34)
R_{merge} (%)	[0.076] 0.126 (1.352)	[0.038] 0.091 (2.306)
R_{meas} (%)	[0.080] 0.131 (1.416)	[0.039] 0.093 (2.424)
R_{pim} (%)	[0.024] 0.036 (0.408)	[0.008] 0.020 (0.740)
$CC_{1/2}$ (%)	[0.998] 0.997 (0.639)	[1.000] 0.999 (0.421)
Mean < I/σ(I) >	[24.3] 10.3 (1.9)	[68.1] 17.5 (1.1)
Completeness (%)	[99.6] 96.9 (79.6)	[100.0] 100.0 (100.0)
No. observed reflections	[4680] 715,684 (25075)	[16955] 2516326 (60014)
No. unique reflections	[4 3 0] 55445 (2169)	[7 9 4] 116165 (5700)
Multiplicity	[10.9] 12.9 (11.6)	[21.4] 21.7 (10.5)
Refinement Statistics		
R_{work}/R_{free}	0.179/0.214	0.144/0.158
RMSD bond lengths (Å)	0.016	0.001
RMSD bond angles (°)	1.819	1.490
Ramachandran statistics (%)		
Favoured	97.74	96.62
Allowed	2.26	3.38
Outliers	0	0
Average B-factors (Å^2)		
Protein	23.5	20.7
Water	33.8	30.8
No. of atoms		
Protein	3319	3504
Water	364	309
PDB Code	**6F0O**	**6F0P**

3. Results and discussion

3.1. Structure of the BoNT/A3 binding domain (H_C/A3)

Crystals of H_C/A3 diffracted to 1.6 Å resolution, and belonged to space group $P2_12_12_1$ with one molecule in the asymmetric unit. A resolution cut-off was imposed based on $CC_{1/2}$ (Evans and Murshudov, 2013). The electron density map obtained following molecular replacement was very good except for three loop regions 1198-1201, 1268-1277 and 1227-1231, which could not be modelled completely. The crystal structure of H_C/A3 contains two motifs common to all BoNTs – a jelly-roll fold consisting of 14 β-strands, and a β-trefoil fold which are predicted to contain a conserved ganglioside and SV2 receptor binding site (Fig. 1a). The H_C/A3 sequence is most similar to that of the BoNT/A2 subtype, where both binding domains share 98.8% identity. This is reflected with a high structural similarity where the RMSD with H_C/A2 is 0.83 Å between Cα atoms.

The ganglioside binding site of BoNT/A1 is mostly conserved in the H_C/A3 structure (Fig. 2i) except for a tyrosine at position 1117 (instead of phenylalanine) and an undetermined conformation of R1276 due to poor electron density. Considering that BoNT/A1 binds to GT1b, it is predicted that BoNT/A3 might do so as well; however, the different residue at position 1117 may affect this since it would no longer be able to form a hydrogen bond to a sialic acid of GT1b (Stenmark et al., 2008).

Recent structures of both H_C/A1 and H_C/A2 bound to their protein receptor, SV2 isoform C (SV2C), suggest that the overall binding mode is generally conserved (Benoit et al., 2014; Yao et al., 2016; Benoit et al., 2017). The high conformational similarity with our structure suggests that H_C/A3 may also bind SV2C in the same manner (Fig. 2b and c). One notable point of variation, however, is position 1152 which is a methionine (and also for subtypes /A4 to /A8) that has the potential to form an electrophilic interaction with H563 of SV2C (Fig. 2c) (Pal and Chakrabarti, 2001). BoNT/A1, however, has an arginine at the equivalent position (Fig. 2a) that forms a cation-π stacking interaction, whereas BoNT/A2 has a glutamate that forms a salt bridge with the equivalent histidine of SV2C. Under normal physiological conditions SV2 is glycosylated, and N-linked glycosylation of SV2C is important for binding of BoNT/A1 to the neuronal cell membrane (Yao et al., 2016; Mahrhold et al., 2016). The glycans form an array of interactions with BoNT residues that neighbour those which interact with the protein chain directly. Within BoNT/A3 (and subtypes /A2, /A6 and /A8), an arginine residue at position 1060 (Fig. 2f), may cause a change to glycan binding specificity compared to BoNT/A1, which has a histidine in the equivalent position (Fig. 2e).

3.2. Structure of the BoNT/A4 binding domain (H_C/A4)

Despite the lower protein concentration of H_C/A4 (due to a lower solubility than H_C/A3) used in the crystal screens, it was possible to grow high quality crystals that diffracted to 1.34 Å resolution, and belonged to space group $P2_13$ with one molecule in the asymmetric unit. The exceptional quality of the electron density map allowed modelling

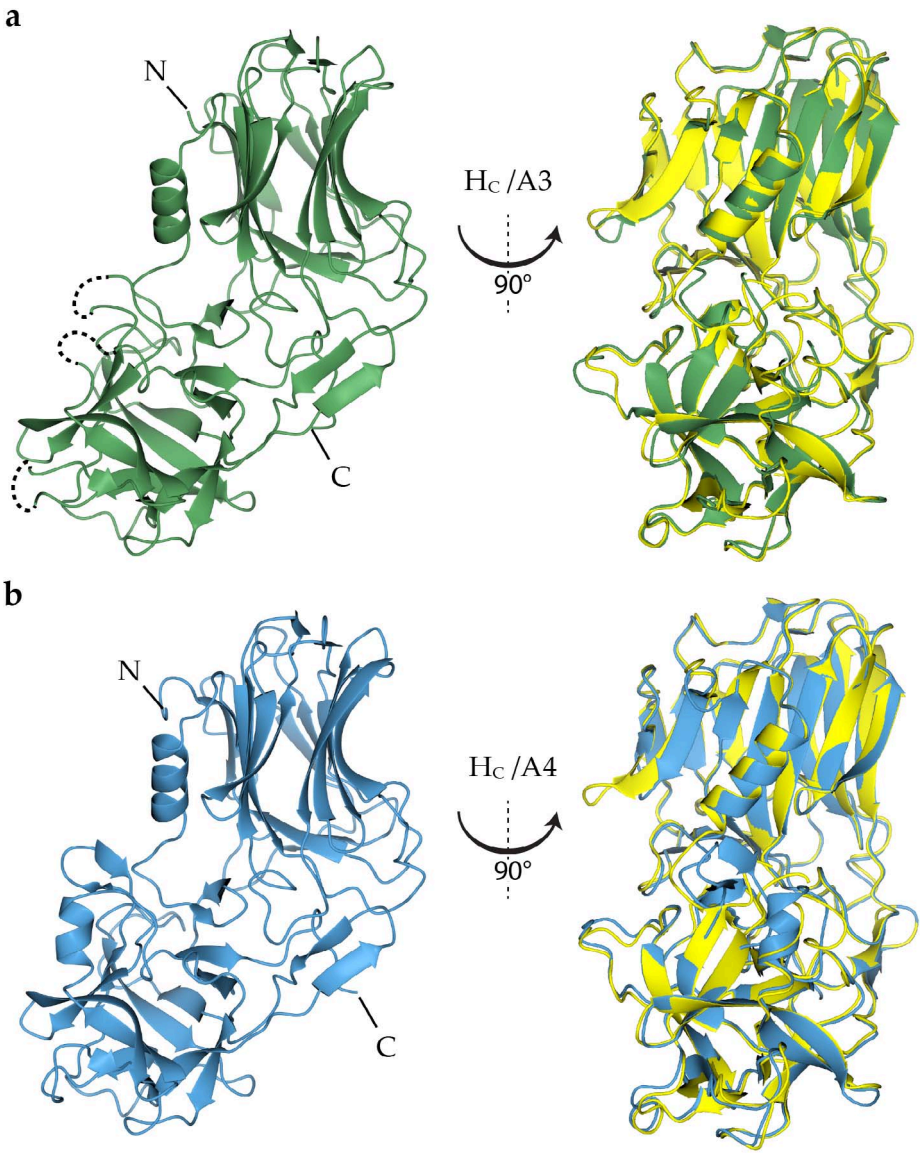


Fig. 1. Structure of the binding domain of BoNT/A3 (**a**) and BoNT/A4 (**b**). Dashed black lines indicate unassigned residues. The BoNT/A1 binding domain (PDB: 2VUA), coloured yellow, is structurally aligned for comparison.

of all residues, as well as identification of a single disulphide bond between C1235 and C1280. As described previously for H_C/A3, the structure of H_C/A4 contains a jelly-roll fold and a β-trefoil fold (Fig. 1b), and all the equivalent residues involved in ganglioside binding for BoNT/A1 (Stenmark et al., 2008) remain unchanged in H_C/A4. This suggests that BoNT/A4 may also share the same ganglioside preference as that of BoNT/A1.

Of all the different subtypes of BoNT/A, the binding domain of BoNT/A4 is most similar to that of BoNT/A1 with 91.78% sequence identity and a RMSD of 0.72 Å between Cα atoms. When comparing binding sites, not only is BoNT/A4 predicted to have the same ganglioside preference as BoNT/A1 (Fig. 2j), it is also predicted to bind to SV2C, according to sequence alignment with BoNT/A1 (Fig. 2a). Most interactions to the SV2C peptide are likely unchanged owing to the nature of backbone-backbone interactions between β-strands, however like H_C/A3, the H_C/A4 structure shows a methionine at position 1156 rather than an arginine at the equivalent position for H_C/A1, which may form an important interaction with SV2C (Fig. 2d). Interactions with SV2C glycans may also be affected by an arginine at position 1292 which protrudes into the binding pocket (Fig. 2g). We, therefore, suggest that, as for BoNT/A3, the binding mode of BoNT/A4 to SV2C itself is likely unchanged when compared to BoNT/A1 but mutations may cause a change to the glycan binding specificity.

4. Conclusion

Subtle structural differences found in both H_C/A3 and H_C/A4 may account for some of the differences in intoxication, such as onset of symptoms and duration of action (Tepp et al., 2012). However, the effect of having different residues within the receptor binding sites, compared to BoNT/A1, has yet to be fully determined. For example, although the variation of glycosylation for recombinant SV2 has been investigated (Mahrhold et al., 2016), the same has not been performed for the native protein receptor at physiological conditions. Differentially glycosylated SV2C may have implications for BoNT receptor targeting and this is one avenue that we are exploring. In addition, the generation of future novel therapeutics relies on the availability of atomic-resolution structures to aid protein engineering, modelling and understanding. Our high-resolution crystal structures, therefore, provide a platform for exploratory work into detailed molecular basis of BoNT binding, as well as structure-based design of novel BoNT therapeutics.

Author contributions

J.R.D. performed all the experiments, analysed the data and wrote the manuscript. J. R. assisted in purification and crystallisation of H_C/

a

Binding Site	BoNT/A1	BoNT/A3	BoNT/A4
SV2 Backbone	S1142	S1138	N1142
	M1144	M1140	M1144
	T1145	T1141	T1145
	T1146	T1142	T1146
	Y1149	Y1145	Y1149
	R1156	M1152	M1156
	R1294	S1290	R1294
SV2 Glycan	F953	F949	F953
	N954	S950	N954
	S957	N953	S957
	D1062	D1058	D1062
	H1064	R1060	H1064
	R1065	R1061	R1065
	T1145	T1141	T1145
	Y1155	Y1151	Y1155
	D1289	D1285	D1289
	G1292	G1288	R1292
	E1293	E1289	E1293
Ganglioside	Y1117	F1113	Y1117
	E1203	E1199	E1203
	F1252	F1248	F1252
	H1253	H1249	H1253
	S1264	S1260	S1264
	W1266	W1262	W1266
	S1275	S1271	S1275
	R1276	R1272	R1276

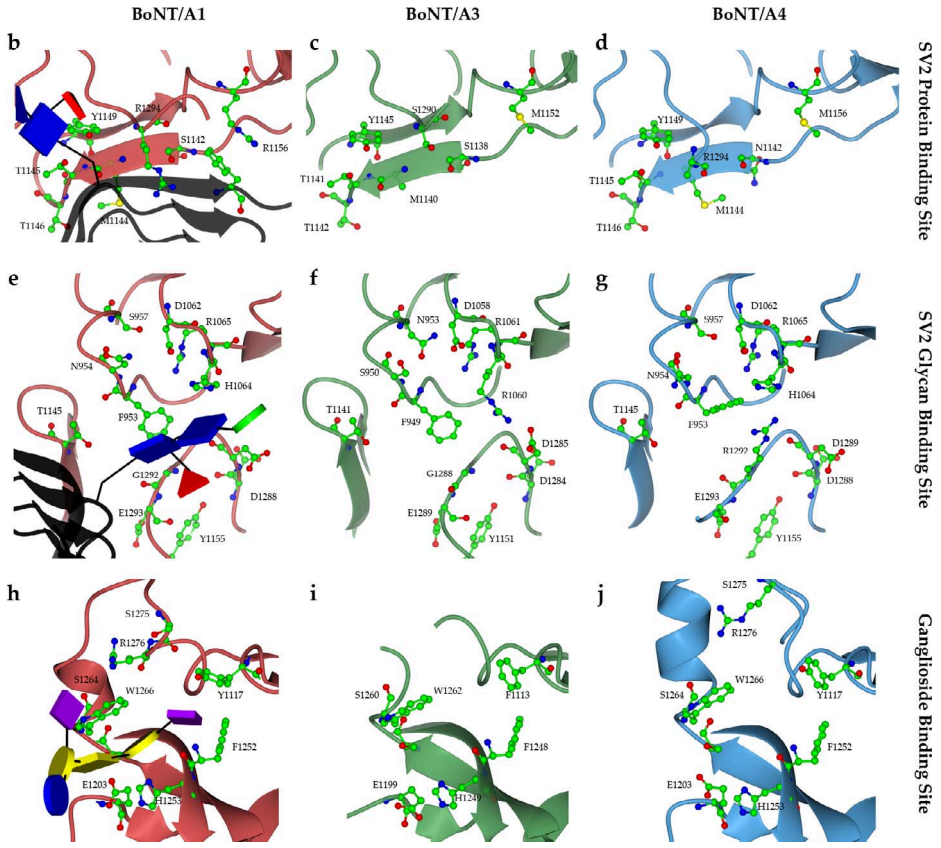


Fig. 2. Comparison of the putative SV2C binding site of BoNT/A3 and BoNT/A4. (**a**) Sequence comparison between residues involved in receptor binding of BoNT/A1, /A3 and /A4 (Genbank: BoNT/A1 WP_011948511.1; BoNT/A3 ABA29017.1; BoNT/A4 ABA29018.1). The predicted SV2C protein binding site of H_C/A3 (**c**) and H_C/A4 (**d**) are conformationally similar to H_C/A1 (**b**), shown here bound to SV2C (in black) with attached glycans (in glycoblock form) (PDB: 5JLV). The glycan binding site of H_C/A1 (**e**) and corresponding residues in H_C/A3 (**f**) and H_C/A4 (**g**), show a similar arrangement. The ganglioside binding site of H_C/A3 (**i**) and H_C/A4 (**j**) are also conformationally similar to H_C/A1 (**h**), shown here bound to ganglioside GT1b (in glycoblock form) (PDB: 2VU9).

A4. S.M.L. analysed the data and edited the manuscript. K.R.A. supervised the study, analysed the data and edited the manuscript. All authors reviewed the manuscript.

Additional information

Accession codes: The atomic coordinates and structure factors (codes 6F0O, 6F0P) have been deposited in the Protein Data Bank (http://www.pdb.org).

Competing financial interests

The authors J.R.D., J.R. and K.R.A. from the University of Bath declare no competing financial interests. S.M.L. is an employee of Ipsen Bioinnovation Limited.

Acknowledgements

We thank Diamond Light Source for access to beamline I03 (proposal mx12342) that contributed to the results presented here. J.R.D. is supported by a joint post-graduate studentship between University of Bath and Ipsen Bioinnovation Limited. This research made use of the Balena High Performance Computing (HPC) Service at the University of Bath.

References

Benoit, R.M., Frey, D., Hilbert, M., Kevenaar, J.T., Wieser, M.M., Stirnimann, C.U., McMillan, D., Ceska, T., Lebon, F., Jaussi, R., Steinmetz, M.O., Schertler, G.F.X., Hoogenraad, C.C., Capitani, G., Kammerer, R.A., 2014. Structural basis for recognition of synaptic vesicle protein 2C by botulinum neurotoxin A. Nature 505, 108–111.

Benoit, R.M., Schärer, M.A., Wieser, M.M., Li, X., Frey, D., Kammerer, R.A., 2017. Crystal structure of the BoNT/A2 receptor-binding domain in complex with the luminal domain of its neuronal receptor SV2C. Sci. Rep. 7, 43588.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murraym, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Cryst. D66, 12–21.

Coffield, J.A., Bakry, N., Carlson, J., Gomella, L.G., Simpson, L.L., 1997. In vitro characterization of botulinum toxin types A, C and D action on human tissues: combined electrophysiologic, pharmacologic and molecular biologic approaches. J. Pharmacol. Exp. Ther. 280, 1489–1498.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Cryst. D66, 486–501.

Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? Acta Cryst. D69, 1204–1214.

Gorrec, F., 2015. The MORPHEUS II protein crystallization screen. Acta Cryst. F71, 831–837.

Henkel, J.S., Jacobson, M., Tepp, W., Pier, C., Johnson, E.A., Barbieri, J.T., 2009. Catalytic properties of botulinum neurotoxin subtypes A3 and A4. Biochemistry 48, 2522–2528.

Johnson, E.A., Montecucco, C., 2008. Botulism Handb. Clin. Neurol. 91, 333–368.

Mahrhold, S., Bergstrom, T., Stern, D., Dorner, B.G., Stot, C., Rummel, A., 2016. Only the complex N559-glycan in the synaptic vesicle glycoprotein 2C mediates high affinity binding to botulinum neurotoxin serotype A1. Biochem. J. 473, 2645–2654.

Mazuet, C., Dano, J., Popoff, M.R., Créminon, C., Volland, H., 2010. Characterization of botulinum neurotoxin type A neutralizing monoclonal antibodies and influence of their half-lives on therapeutic activity. PLoS One 5, e12416.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. J. Appl. Cryst. 40, 658–674.

McNicholas, S., Potterton, E., Wilson, K.S., Noble, M.E.M., 2011. Presenting your structures: the CCP4mg molecular-graphics software. Acta Cryst. D67, 386–394.

Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst. D53, 240–255.

Rossetto, O., Pirazzini, M., Montecucco, C., 2014. Botulinum neurotoxins: genetic, structural and mechanistic insights. Nat. Rev. Microbiol. 12, 535–549.

Pal, D., Chakrabarti, P., 2001. Non-hydrogen bond interactions involving the methionine sulfur atom. J. Biomol. Struct. Dyn. 19, 115–128.

Schiavo, G., Matteoli, M., Montecucco, C., 2000. Neurotoxins affecting neuroexocytosis. Physiol. Rev. 80, 717–766.
Stenmark, P., Dupuy, J., Imamura, A., Kiso, M., Stevens, R.C., 2008. Crystal structure of botulinum neurotoxin type A in complex with the cell surface co-receptor GT1b—insight into the toxin–neuron interaction. PLoS Pathog. 4, e1000129.
Tepp, W.H., Lin, G., Johnson, E.A., 2012. Purification and characterization of a novel subtype A3 botulinum neurotoxin. Appl. Environ. Microbiol. 78, 3108–3113.
Waterman, D.G., Winter, G., Gildea, R.J., Parkhurst, J.M., Brewster, A.S., Sauter, N.K., Evans, G., 2016. Diffraction-geometry refinement in the DIALS framework. Acta Cryst. D72, 558–575.
Whitemarsh, R.C., Tepp, W.H., Johnson, E.A., Pellett, S., 2014. Persistence of botulinum neurotoxin A subtypes 1–5 in primary rat spinal cord cells. PLoS One 9, e90252.
Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., 2011. Overview of the CCP4 suite and current developments. Acta Cryst. D67, 235–242.
Yao, G., Zhang, S., Mahrhold, S., Lam, K.H., Stern, D., Bagramyan, K., Perry, K., Kalkum, M., Rummel, A., Dong, M., Jin, R., 2016. N-linked glycosylation of SV2 is required for binding and uptake of botulinum neurotoxin A. Nat. Struct. Mol. Biol. 23, 656–662.

SCIENTIFIC REPORTS

OPEN

Structural analysis of *Clostridium botulinum* neurotoxin type D as a platform for the development of targeted secretion inhibitors

Received: 08 June 2015
Accepted: 27 July 2015
Published: 01 September 2015

Geoffrey Masuyer[1,3], Jonathan R. Davies[1], Kevin Moore[2], John A. Chaddock[2] & K. Ravi Acharya[1]

The botulinum neurotoxin type D is one of seven highly potent toxins produced by *Clostridium botulinum* which inhibit neurotransmission at cholinergic nerve terminals. A functional fragment derived from the toxin, LHn, consisting of the catalytic and translocation domains, has been heralded as a platform for the development of targeted secretion inhibitors. These secretion inhibitors are aimed at retargeting the toxin towards a specific cell type to inhibit vesicular secretion. Here we report crystal structures of LHn from serotype D at 2.3 Å, and that of SXN101959 at 3.1 Å resolution. SXN101959, a derivative that combines LHn from serotype D with a fragment of the growth hormone releasing hormone, has previously revealed promising results in inhibiting growth hormone release in pituitary somatotrophs. These structures offer for the first time insights into the translocation domain interaction with the catalytic domain in serotype D. Furthermore, structural information from small-angle X-ray scattering of LHn/D is compared among serotypes A, B, and D. Taken together, these results demonstrate the robustness of the 'LHn fold' across serotypes and its use in engineering additional polypeptide components with added functionality. Our study demonstrates the suitability of botulinum neurotoxin, and serotype D in particular, as a basis for engineering novel secretion inhibitors.

Toxins from *Clostridium botulinum* species are the causative agent of the rare neuroparalytic illness botulism. Seven distinct serotypes (A–G) of botulinum neurotoxins (BoNTs) affect humans and other species to varying degrees. Once inside the neuronal cell, BoNTs block the release of neurotransmitters leading to paralysis[1]. Although they are highly toxic, various BoNTs are available commercially as therapeutic agents[2].

BoNTs are synthesized as a single polypeptide chain (150 kDa), which is post-translationally cleaved into a di-chain molecule composed of light chain (LC, 50 kDa) and heavy chain (HC, 100 kDa). LC is the catalytic domain and a zinc-endopeptidase, while HC is further divided into two sub-domains of equal molecular mass called the translocation domain (Hn) and the membrane binding domain (Hc). On binding to the nerve terminals, BoNTs are endocytosed into a vesicle, where the acidic environment causes some conformational changes allowing LC to enter the cytosol[3]. Inhibition of neurotransmission takes place by proteolysis of one of the SNARE proteins[1] that mediate cell secretion[4].

Of the seven types of botulinum toxin, A, B, E and F are known to cause the disease in human while C and D have only been observed in animal cases. More particularly, D has been responsible for several

[1]Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK. [2]Ipsen Bioinnovation Limited, Units 4-10, The Quadrant, Barton Lane, Abingdon, Oxon OX14 3YS, UK. [3]Current address: Department of Biochemistry and Biophysics, Arrhenius Laboratories for Natural Sciences, Stockholm University, 10691 Stockholm, Sweden. Correspondence and requests for materials should be addressed to K.R.A. (email: bsskra@bath.ac.uk)

recent outbreaks of botulism in cattle[5]. This has raised some interest in this serotype and understanding its precise mechanism of action. No cases of type D human botulism have ever been recorded. Coffield *et al.*[6] investigated the impact of serotype D on human tissues and demonstrated its inability to block neuromuscular transmission when tested at a level 10 times higher than that of serotype A. This limited activity was recently confirmed *in vivo* by electrophysiological study on human muscles[7] and may be linked to differences in the receptor binding domain[8,9]. BoNT/D however, acts similarly to the other botulinum neurotoxins by targeting one of the intracellular SNARE proteins. Synaptobrevin (or VAMP) was identified as the BoNT/D specific substrate[10] and is cleaved at the Lys59-Leu60 position.

BoNT/A and /B are currently the only serotypes approved for medical uses. With the emergence of immuno-reactivity and resistance among patients[11], other serotypes such as BoNT/D can provide a useful alternative. The successful retargeting of BoNT activity for therapeutic purposes, by association of the LHn fragment (catalytic and translocation domains) with various ligands has been described[12]. In particular, a targeted secretion inhibitor (TSI) combining a growth hormone releasing hormone (GHRH) receptor targeting domain with the LHn/D fragment, called qGHRH-LHn/D, was shown to specifically inhibit pituitary somatotroph growth hormone release[13]. This molecule demonstrated efficient intracellular activity on VAMP-3 in rat pituicytes and thus encouraging potential in the treatment of Acromegaly[14]. The latest LHn/D derivative, SXN101959, which combines the functional BoNT fragment with a GHRH ligand domain [qGHRH(1–40)], presented potent and reversible inhibitory action on the somatotropic axis and such features are well aligned with treating overproduction of growth hormone[15]. The arrangement of the functional domains within SXN101959 involves a novel orientation in which the GHRH ligand is located centrally between the LC and HC in the single chain polypeptide expressed in *E. coli* and with a protease cleavage site located between the GHRH ligand and the LC, such that following protease activation to generate the active di-chain TSI the GHRH ligand is at the amino-terminus of the HC domain. This arrangement, termed central presentation, was necessitated by the requirement of the GHRH ligand to have a free N-terminus to be able to bind and activate its receptor. It means, however, that the relative arrangement of the binding domain relative to the LC and Hn domains is different to that in BoNT and previously reported TSI proteins[16]. The resulting TSI protein is functional in respect to binding its receptor, internalising into the cytosol of target cells, cleaving its substrate SNARE protein and inhibition of growth hormone secretion[14,15]. However, the consequence of this novel domain arrangement upon the structure of the TSI protein is not known.

Structural and biochemical characterisation of the LHn fragments from serotypes A and B[17,18] have provided a molecular basis for their functionality, representing an important step forward for the design of novel molecules based on these frameworks. In order to assess its applicability for further pharmaceutical development, the structure of LHn/D was analysed by X-ray crystallography (solid crystalline state) and small angle X-ray scattering (SAXS, in the solution state). The X-ray crystal structure showed key elements of BoNT/D including the flexibility of the translocation domain and the elaborate interaction with its catalytic partner. The SAXS analysis corroborates the relevance of the crystal structures and the stability of the LHn template in solution. The X-ray crystal structure of SXN101959 at 3.1 Å further confirmed the rigidity of LHn/D's structural framework and its ability to accommodate novel arrangements of the domain subunits such as central presentation, although the GHRH ligand domain was not visible. This study should thus provide the structural basis for the development of BoNT/D-based Targeted Secretory Inhibitor (TSI) which could prove useful in the treatment of hypersecretory disorders.

Results and Discussion

Structure of LHn/D. This is a mutant endo-negative form of LHn/D in which the catalytic domain is inactivated by a two amino acid substitution in the zinc-binding site (H233Y, E230Q). The crystal of LHn/D diffracted at 2.3 Å resolution in space group $P6_422$ with cell dimensions $a = b = 173$, $c = 222$ Å; $\alpha = \beta = 90$, $\gamma = 120°$ (Table 1). Despite the large unit cell, a single molecule was present in the asymmetric unit corresponding to an unusually high solvent content of 75%. The X-ray crystal structure presents the two domains, LC and Hn, in a close interaction. No electron density was observed for regions 442–457 (between the two domains), residues 494–510, and the C-terminal poly-histidine tag (last fifteen residues), all in solvent-accessible areas (Fig. 1).

LHn/D share 34 to 35% identity with the corresponding fragment of BoNT/A, /B and, /E, for which the full length structures are available[19,20,21]. LHn/D presents a similar overall fold as previously observed for LHn/A and /B[17,18]. The translocation domain of serotype D is shown for the first time and includes two long coiled-coil helices with additional short helical segments on both sides. The catalytic domain has the globular fold typical of other light chains but shows key differences in flexible loop regions. Additionally, one of the most intriguing features of BoNT, the 'belt' region (residues 458–547), which is part of Hn and surrounds LC, was visible in the electron density map (despite high solvent content in the crystal) in the structure presented here and highlights the intricate interactions between the two domains. A PISA[22] analysis shows that the LC-belt interaction is supported by over 40 potential hydrogen bonds and 2 salt bridges, making up a 3373 Å^2 contact area. Additionally LC and Hn (excluding the belt) make a further 1392 Å^2 buried interface area principally involving hydrophobic interactions.

The light chain. The structure of the catalytically inactive mutant light chain superposes well with the high-resolution structure of a recombinant form previously reported (PDB code 2FPQ[23]) with a root

	LHn/D	SXN101959
Resolution (Å)[a]	2.3 (2.3–2.34)	3.1 (3.1–3.25)
Space group	$P6_422$	$P2_12_12_1$
Cell dimensions (Å; a, b, c); angle (°; α, β, γ)	173, 173, 222; 90, 90, 120	88, 144, 173; 90, 90, 90
Total/Unique reflections	321,682/86,306	122,006/35,774
Completeness (%)[a]	98.8 (97.7)	89.0 (81.2)
R_{merge}[a,b]	0.059 (1.09)	0.127 (1.02)
R_{pim}[a,c]	0.048 (1.04)	0.101 (0.859)
I/σ(I)[a]	8.1 (0.5)	6.0 (1.0)
$CC_{1/2}$[d]	0.998 (0.214)	0.990 (0.331)
Multiplicity	3.7 (2.6)	3.4 (2.6)
R_{cryst}[e]	22.0%	25.0%
R_{free}[f]	24.7%	29.5%
Rmsd in bond lengths (Å)	0.011	0.007
Rmsd in bond angles (°)	1.45	0.97
B- factor statistics (Å^2)		
Protein all atoms	79.1	94.6/110.3[g]
Protein main chain atoms	78.3	95.0/110.7
Protein side chain atoms	79.9	94.2/109.8
Zinc ion	N/A	64.9/85.4
Solvent atoms	66.4	51.0
Ramachandran statistics (Molprobity)		
Favoured	96.2%	93.9%
Outliers	0.2%	0.5%
PDB code	**5BQN**	**5BQM**

Table 1. X-ray crystallography: data collection and refinement statistics. [a]Values in parentheses refer to the highest resolution shell. [b]$R_{merge} = \Sigma\Sigma_i|I_h - I_{hi}|/\Sigma\Sigma_i I_h$, where I_h is the mean intensity for reflection h. [c]$R_{pim} = \Sigma_h (1/n_h - 1)\ \Sigma_l |I_{hl} - (I_h)|/\Sigma_h\Sigma_l(I_h)$. [d]Correlation coefficient between random half datasets[35]. [e]$R_{cryst} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are measured and calculated structure factors, respectively. [f]$R_{free} = \Sigma||F_o| - |F_c|/\Sigma|F_o|$, calculated from 5% of the reflections selected randomly and omitted during refinement. [g]The two B-factor values recorded here correspond to two molecules in the asymmetric unit.

mean square deviation (rmsd) of 2.3 Å (over 411 Cα atoms; Fig. 2). Loss of the zinc ion was confirmed with a local rearrangement of the active site residues (Fig. 2B,C). Mutation of the zinc-coordinating H233 to tyrosine caused displacement of the catalytic ion, replaced in its position with the Y233 hydroxyl group. Overall, the role of the zinc ion appears mostly catalytic with little impact on the structural integrity of the light chain.

The 2FPQ[23] structure (hereafter referred to as '2FPQ') corresponds to residues 1–424 of the light chain and thus do not include some of the elements interacting with the rest of the toxin. In this structure, residues 166–170 of the catalytic channel appear in two possible conformations, one is a random coil close to the zinc atom and the second a short β-strand completing the catalytic pocket. Furthermore, residues 173–179 were disordered. The present LHn/D structure, in contrast, was stable across residues 166 to 179 where segments 168–172 and 178–184 take on an α-helical arrangement (Fig. 2D). The multiple conformations of this area across the crystal structures, in vicinity of the active site, indicate a degree of flexibility that may provide the necessary space for substrate interaction[24].

The most noticeable differences between the new structure and 2FPQ reside in flexible loop regions, particularly those at the interface with the translocation domain. For example, loop 63–70, which is surface accessible in 2FPQ, still presents a random coil in LHn/D but is stabilised by the extended carboxyl-terminal end of LC (residues 429–431) on one side (S70) and by the end of the 'belt' (F539) on the other (R63) (Fig. 2D). Additionally, loop 250, which presents low sequence identity with other light chains[23] but is visible in 2FPQ, adopts a different conformation in LHn/D with a 10 Å and 25° movement (Fig. 2D). The loop is stabilised through hydrophobic interactions involving residues 256–258 (GFF) with Hn, and potential electrostatic interactions of residues Q260-D261 with the belt which occludes the catalytic pocket. A third area of interest is the flexible loop at positions 205–217 which was completely disordered in 2FPQ and is visible in LHn/D. The hydrophobic part of the loop (V205-T206) fits within the aromatic pocket composed of Y731 and Y734 of one of the long Hn helices, while the backbone of

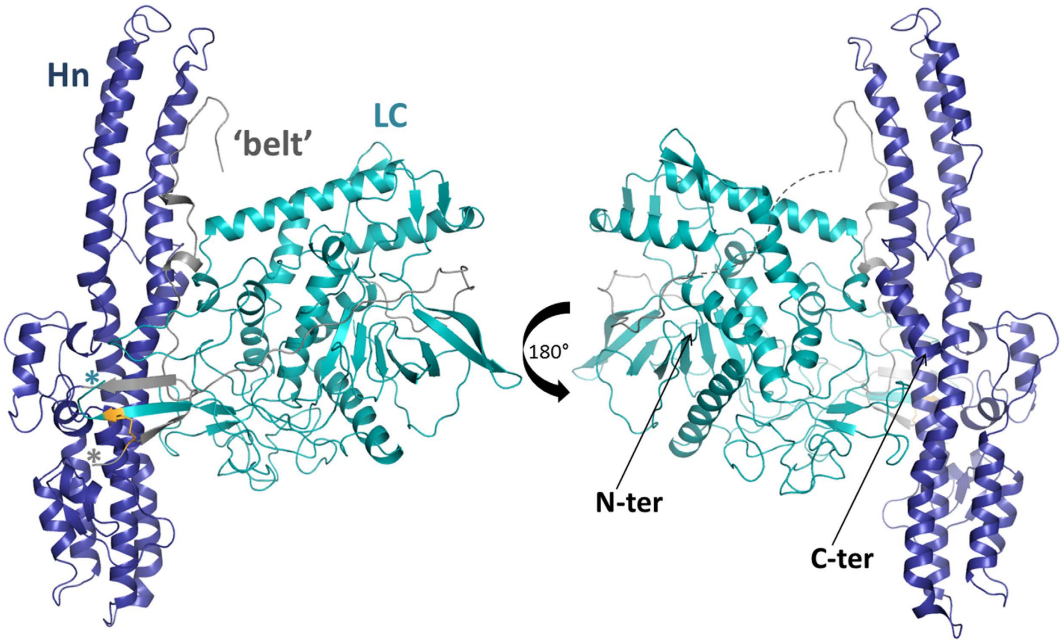


Figure 1. Crystal structure of LHn/D. Ribbon diagram representation of LHn/D structure, LC in cyan, belt region in grey and Hn in blue. C-terminus of LC and N-terminus of Hn are marked with a cyan and grey asterisk, respectively. The disordered region of the belt is presented as a grey-dashed line. The disulphide bond between LC and Hn is highlighted in orange.

residues 212–214 is stabilised by hydrophobic interactions with the other helix (V776-S779) (Fig. 2E). Although crystal packing might influence some of these interactions, the observed changes could partly be attributed to the complex interface between the two domains compared to the structure of the single light chain (in 2FPQ).

The interaction between the catalytic domain and its VAMP substrate involves several binding sites based on the information provided by the structure of LC/F with a VAMP-inhibitor complex[25] and recent mutagenesis work[24]. Several sub-sites of the catalytic pocket of LC/D have been shown to be essential for substrate binding. Interestingly the LHn/D structure highlights the plasticity of LC within the context of the whole toxin at these particular sites. For example R372 of S2′ (Fig. 2D) presents a double conformation but is expected to form a hydrogen bond with the P2′ site of VAMP. Additionally, the hydrophobic S1′ pocket and the S3 loop of the catalytic domain in LHn/D present new alternative conformations which were shown to cause at least a 50-fold loss of catalytic activity when mutated to Y168A/L200D and R63A, respectively (Fig. 2D)[24].

The long substrate requirement of all BoNT serotypes is atypical to other zinc proteases and involves several binding sites distal from the catalytic pocket[26]. Guo and Chen identified and labelled these exosites B1 to B5 in LC/D[24]. None of these sites present any conformational changes in LHn/D. While B1 is buried within LC, 15 Å away from the active site, the hydrophobic patches B4 to B5 show the extent of substrate recognition at the accessible surface of the free light chain. In LHn/D these pockets are interacting with the belt (Fig. 3) and implies a potential substrate inhibiting role for this region as also hypothesised for BoNT/A and /B[27].

The translocation domain and the belt. The catalytic domain is normally linked to the translocation domain through a single disulphide bridge after activation of BoNT into a di-chain molecule. Although the construct used for crystallisation experiment was not activated, residues 442–457, corresponding to the engineered cleavage site and flexible linker between the two domains, were not visible. The cysteine bridge is however stabilised by formation of a small β-sheet between the C-terminus of LC, the N-terminus of Hn and the end of the belt segment (Fig. 1).

The belt region in /D is overall similar to the corresponding segments in serotypes /A, /B and /E (Fig. 3). It surrounds LC following a hydrophobic groove along its surface. Interestingly, the belt in the other known BoNT structures has been observed in its entirety and takes on a conserved α-helical arrangement at a position where LC/A was shown to bind its substrate (SNAP-25) in a similar secondary conformation[21,26]. In LHn/D, seventeen amino acids of this area are missing altogether (residues 494–510), indicative of a flexible loop region which is unlikely to interact with LC/D. Remarkably, LC/D is known to recognise a shorter substrate than its counterparts[28], and this site is therefore not expected to play a role in VAMP binding. The full function of the belt region in BoNT is not yet completely

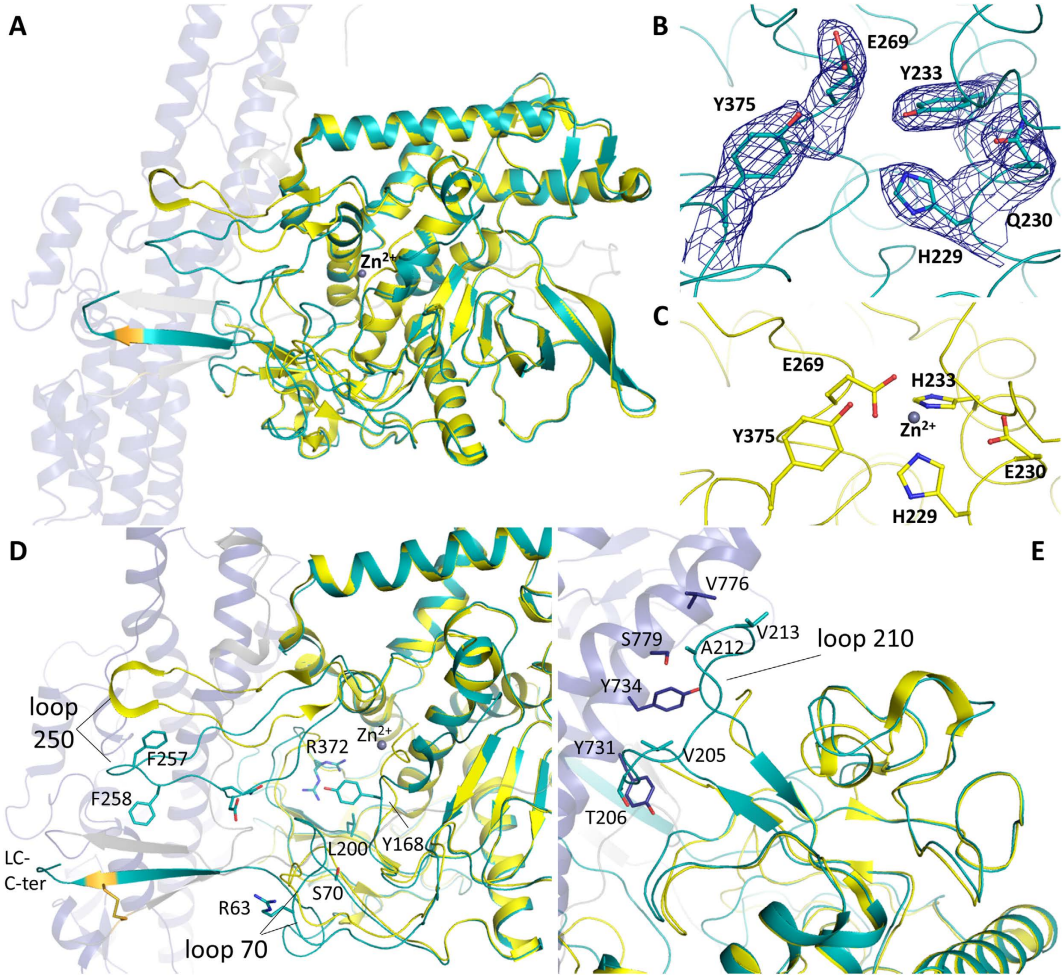


Figure 2. Comparison of the LC structure of LHn/D with LC/D. (**A**) Ribbon diagram representation of LHn/D structure with its light chain in cyan, belt region in grey and translocation domain in blue. PDB 2FPQ[23] is shown in yellow with the zinc ion as a grey sphere. (**B**) Residues of the catalytic pocket of LHn/D are shown with sticks to highlight the effect of the double mutation E230Q and H233Y. The *2Fo-Fc* electron density map countered at 1σ level is shown in blue. (**C**) Residues of the catalytic pocket of LC/D presenting the classic zinc coordination motif HEXXH. (**D**) and (**E**) Close-up views of loop 250 and 210, respectively. Residues involved in loop interactions are shown in sticks and labelled.

understood but evidence suggests a possible role beyond protection of LC, and towards a chaperone-like mechanism for translocation and membrane insertion[27,29].

The main components of the translocation domain are the two coiled-coil helices of approximately 105 Å (Fig. 1). While the secondary α-helical structure is conserved across serotypes, the two helices all have a central bending point which curves to various degrees (Fig. 4). Of the four serotypes (with known Hn structures thus far), D appears the straightest. It should also be noticed that Hn/D presented particularly high B-factor at the extremities of these helices (Fig. 5D), with weak electron density for the random-coiled loop linking them. This is likely an indication of flexibility within these regions. The smaller helices flanking the main frame are also well conserved and similarly positioned across serotypes. Additionally the visible C-terminal end forms a small α-helix which is involved in crystal packing with a symmetry-related molecule and appears in a different orientation to the similar segment in LHn/A and /B.

Structure of SXN101959. SXN101959 differs from LHn/D, and from natural BoNT molecules at the linker region between the two domains. It includes a Factor Xa cleavage site, a GHRH ligand domain (qGHRH[1–40]) and a GGGGS repeat (totalling an additional 56 residues). The engineered protein was crystallised as an activated di-chain molecule and diffracted up to 3.1 Å resolution (Table 1). The two molecules of the asymmetric unit are overall similar (rmsd of 1.05 Å over 828 Cα atoms) and present the same disordered regions which correspond to 66 amino acids of the linker peptide (described above), and 17 residues of the belt region. Both these areas were also not visible in LHn/D (Fig. 5).

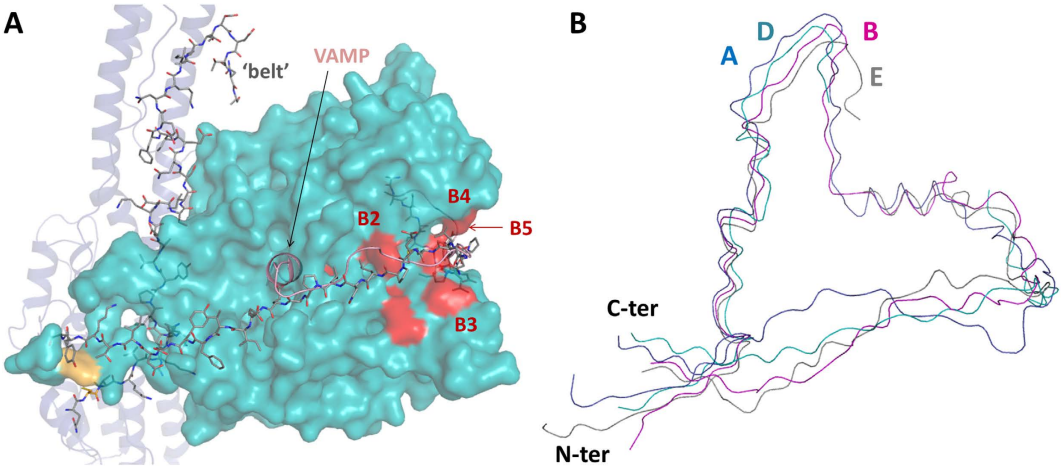


Figure 3. The belt of BoNT/D. (**A**) Crystal structure of LHn/D with the belt region represented with grey sticks. The rest of the translocation domain is shown in blue, the light chain in shown as a cyan surface. The structure of LC/F in complex with a VAMP peptide inhibitor (PDB 3FIE[25]) was superposed to LHn/D and VAMP is represented as a pink ribbon. (**B**) Superposition of the belt regions of LHn/A (PDB 2W2D[17]), /B (PDB 2XHL[18]), /D and BoNT/E (PDB 3FFZ[21]), in blue, magenta, cyan and grey ribbon, respectively. The amino- and carboxyl-end of the belt are labelled.

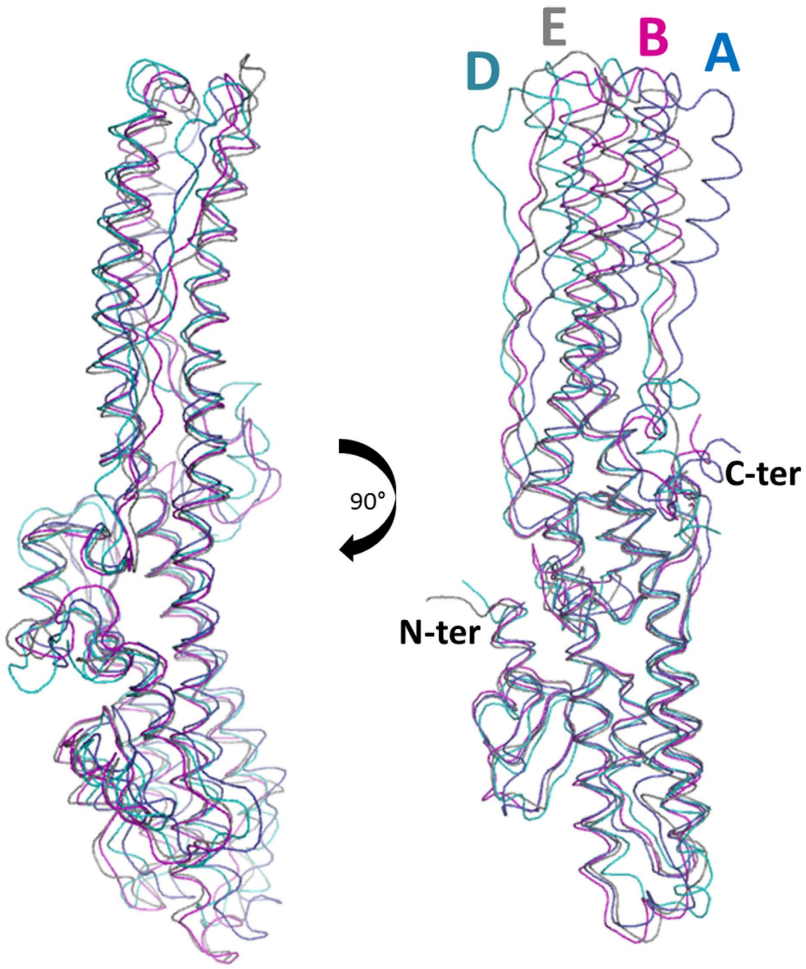


Figure 4. Comparison of BoNT translocation domains. Superposition of the translocation domains (without the belt) of LHn/A, /B, /D and BoNT/E, in blue, magenta, cyan and grey ribbon, respectively. The amino- and carboxyl-end of the domain are labelled.

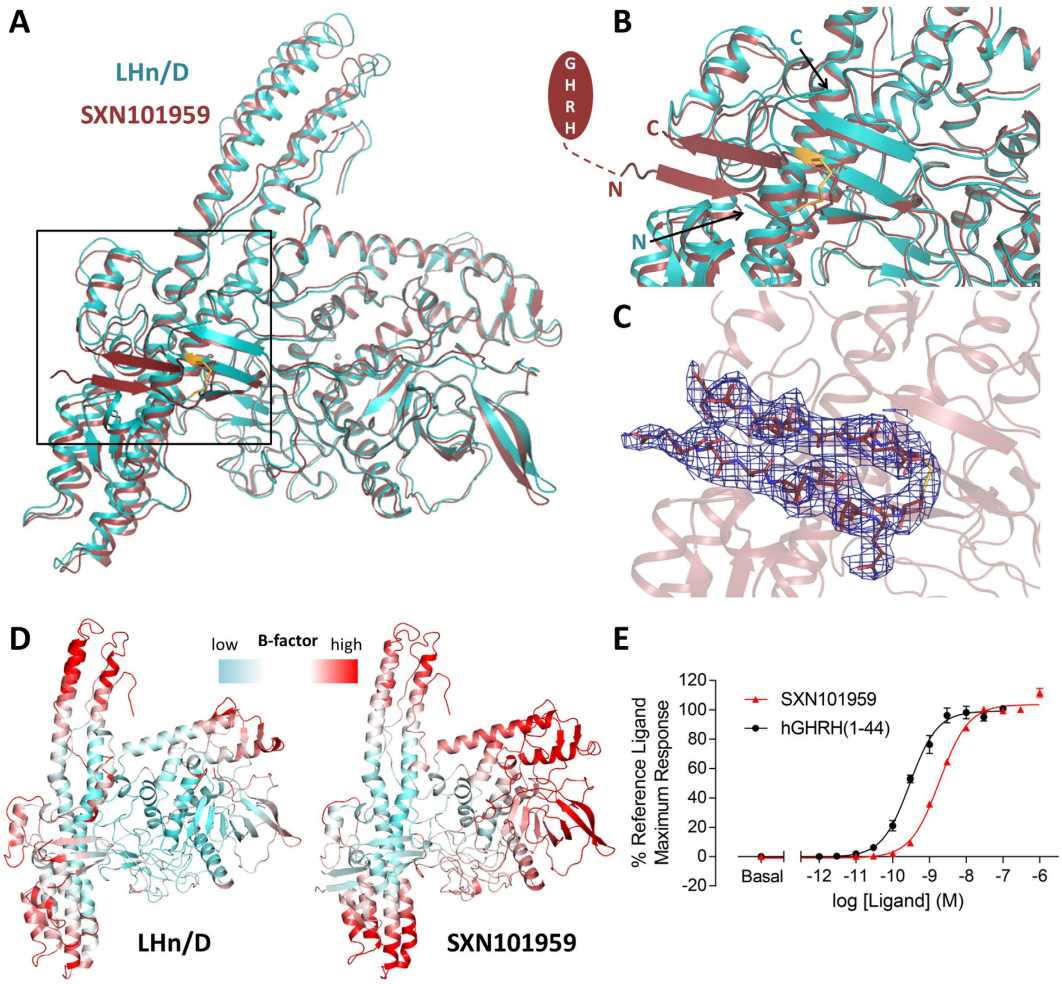


Figure 5. Crystal structure of SXN101959 and comparison with LHn/D. (**A**) The crystal structure of SXN101959, in dark red, was superposed with LHn/D, in cyan. The disulphide bridge between LC and Hn is highlighted in orange. The zinc ion of SXN101959 is shown as a grey sphere. (**B**) Close-up view of the domain interface in SXN101959 and LHn/D. The C-terminal end of LC and N-terminus of Hn are labelled. The location of the targeting ligand is represented with the missing part of the linker shown as a dashed line. (**C**) The visible part of the linker region of SXN101959 is highlighted with the *2Fo-Fc* electron density map countered at 1σ level shown in blue. (**D**) Temperature factor (B-factor) analysis. Ribbon diagram representation of LHn/D and SXN101959 structures, with a gradient colouring from low (cyan) to high (red) B-factors. Gradients were adjusted independently to highlight the less ordered area for each structure. (**E**) **Activation of the human GHRH receptor**. The receptor activation was detected by the intracellular accumulation of cAMP in CHO-K1-hGHRH-R cells incubated with hGHRH(1–44) (●) or SXN101959 (▲). Data are mean ± sem of triplicate samples from one experiment.

Although the targeting ligand could not be observed, the crystal structure still provides valuable information on LHn/D as a framework for protein engineering. Indeed addition of a large peptide segment at the LC-Hn interface, which is essential for BoNT activity[3], did not affect the LHn/D scaffold (each molecule superposes to LHn/D with rmsd of 1.08 (821 Cα) and 1.40 Å (828 Cα atoms), respectively). The light chains of SXN101959 and LHn/D are identical, with little variation in the solvent accessible loop regions. No noticeable difference was seen in the backbone of the active site (Fig. 5) despite SXN101959 being catalytically active. Clear electron density provided evidence of the zinc ion at the catalytic site, which is coordinated as per the classical HEXXH motif. Unsurprisingly, the more noticeable difference between SXN101959 and LHn/D resides in the translocation domain where the extremities of the coiled-coil helices are showing signs of flexibility, confirming what was observed in LHn/D (Fig. 5).

The linker region (residues 438–513 in SXN101959) was engineered to provide ease of access of the targeting GHRH ligand for its receptor. It is flanked with an N-terminal protease cleavage site that allows activation of the whole molecule into its di-chain form, and frees the amino end of the peptide chain. The biological activity of GHRH is retained by the 29 N-terminal amino acids for recognition by the

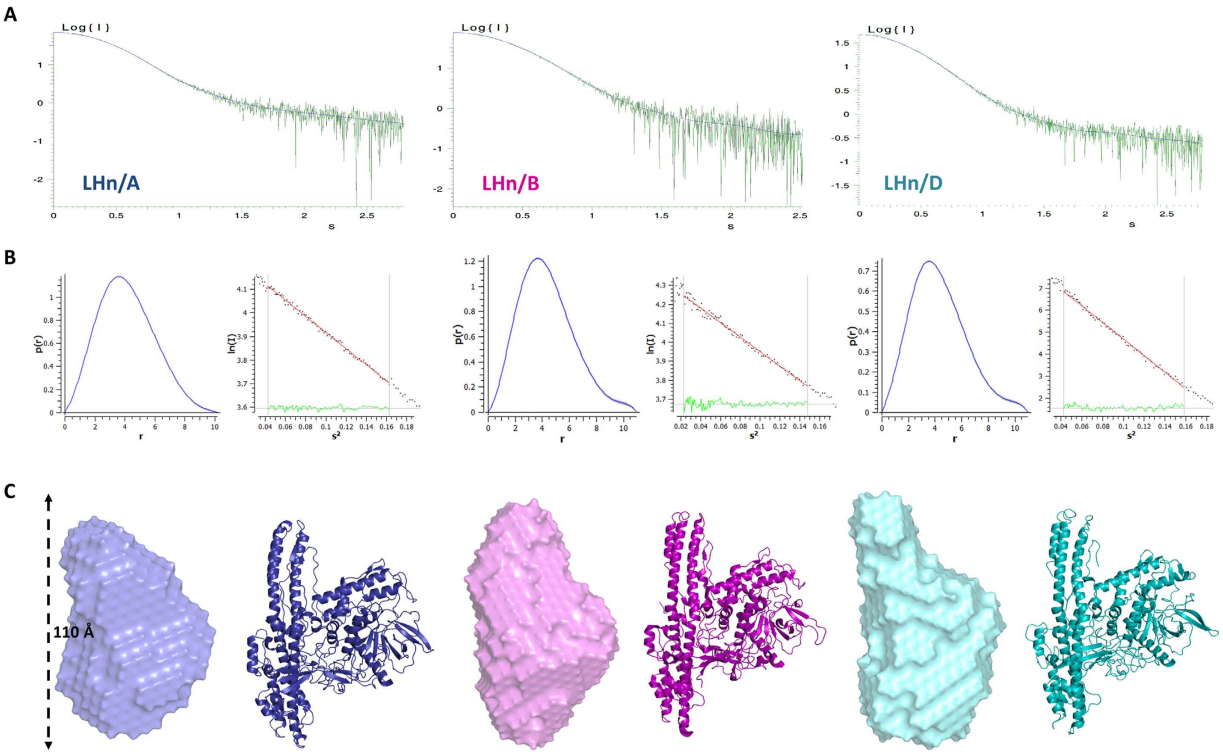


Figure 6. Small-angle X-ray scattering (SAXS) on LHn. (**A**) Scattering curve of LHn/A, /B and /D, from left to right. (*I* is a relative unit, *s* is expressed in nm^{-1}). (**B**) The interatomic distance function (*P(r)*) was calculated (x axis, r in nm) with GNOM[43] (left) and Guinier plot is presented for each samples. (**C**) *Ab initio* solution scattering models. SAXS bead models were generated with DAMMIF[44] and averaged in DAMAVER[45]. The models are compared to the samples' respective X-ray crystal structures.

GHRH receptor[30]. In order to give the system more structural flexibility, three GGGGS repeat motifs were included on the C-terminal side of the peptide. This allows for the targeting peptide to be optimally and centrally presented while it remains covalently linked to the translocation domain. This flexible segment, was disordered in the crystal structure. It is located in a solvent accessible area where it could be easily accommodated within the crystal packing (Supplementary Figure). It should also be noticed that the observable part of the linker, consisting of the LC's carboxyl and Hn's amino ends (438–443 and 509–513, respectively), is stabilised through the formation of an anti-parallel β-sheet conformation (Fig. 5) so that one of the GGGGS repeats is visible. The targeting peptide is not interacting with the rest of the molecule, which is encouraging from a biological perspective, as it relies on this freedom of movement to activate the GHRH receptor. SXN101959 is known to activate GHRH receptor in rat primary pituicytes which leads to internalisation and intracellular cleavage of VAMP3[14].

Integrity of the sample used for structural study was tested by measuring the intracellular accumulation of the second messenger cAMP in human GHRH receptor-expressing cells. SXN101959 elicited a concentration-dependent increase in cAMP accumulation with a pEC50 (potency) of 8.73, compared with the pEC50 of hGHRH(1–44) of 9.54. The maximum response of SXN101959 was the same as hGHRH(1–44), indicating that the TSI is a full agonist of hGHRH receptor. Activation of the target receptor confirms the presence of the qGHRH(1–40) ligand within the structure of SXN101959 (Fig. 5E). Thus the crystal structure of SXN101959 highlights the validity of LHn/D in the context of a biologically active TSI.

Small-angle X-ray scattering studies on LHn. The botulinum neurotoxin has been extensively studied by X-ray crystallography but little information is available on the relevance of these structures in solution. Small-angle X-ray scattering (SAXS) experiments were carried out on LHn/A, /B and /D. All data were collected on station X33 at DESY (Germany) and the solution scattering curves are provided in Fig. 6A. The purified samples were all in their activated di-chain form.

A Guinier analysis of the scattering curves was consistent with the distance distribution function (*P(r)*) in determining the radius of gyration (R_g) for each LHn serotypes with values of approximately 32, 34 and 34 Å for A, B and D, respectively (Table 2; Fig. 6B). Interestingly, LHn/B and /D also share a similar maximum dimension (D_{max}) of 110 Å, slightly longer than LHn/A at 102 Å. The comparable

	Guinier[1]	*P(r)* function[1]		Crystal structures[2]		Structural modeling[3]	
	R_g (Å)	R_g (Å)	D_{max} (Å)	R_g (Å)	Fit (χ^2)	Discrepancy (χ^2)	NSD
LHn/A	32.1	32.0	102	31.3	1.5	0.88	0.60
LHn/B	33.9	33.9	110	32.3	1.0	0.81	0.60
LHn/D	33.7	33.7	110	31.4	2.0	0.97	0.62

Table 2. Small angle X-ray scattering data statistics for LHn serotypes /A, /B and /D. [1]The Guinier analysis and distance distribution function (*P(r)*) were performed with PRIMUS[42] and used to calculate the maximum dimension (D_{max}) and the radius of gyration (R_g). [2]The X-ray crystal structures of LHn/A, /B and /D were used for theoretical scattering calculation and compared to experimental data using CRYSOL[43]. [3]The discrepancy (χ2) between experimental and *ab initio* data was calculated by the program DAMMIF[44] with the average of 10 models (NSD = normalized spatial discrepancy).

features of the three samples correlate with the known crystal structures. The theoretical scattering curves extrapolated from the high-resolution structures were calculated with CRYSOL[31] and fitted well with the experimental values (Table 2) although the corresponding R_g were somewhat lower than the experimental ones (between 31 and 32 Å). *Ab initio* models were generated from the scattering curves (Fig. 6C, Table 2), and present the overall shape of LHn in solution.

Manual docking allowed the comparison between the *ab initio* bead models and the crystal structures. The distinctive shape of the two domains forming the LHn molecules can be clearly noticed in all three samples. Indeed the atypical long coiled-coil helices of the translocation domain give the models an extended form over 100 Å long whereas the light chain accounts for the globular central shape. In accordance with the statistical comparison (Table 2), LHn/B appears to show the best resemblance between crystal and solution structures. The smaller dimensions of LHn/A in solution might be the reflection of the more pronounced central kink of its translocation domain (Fig. 4). Additionally, LHn/D presents a less defined upper region which might be an indication of domain movement, considering the bead model represents an averaged form of the solution state. Interestingly the extremities of the translocation domain were also showing signs of mobility in the crystal structures of LHn/D and SXN101959.

Overall, the information collected for LHn/A, /B and /D showed that the fragments have conserved the LHn fold in the solution state although the resolution of the data may not allow the visualisation of small variations.

Conclusion

Botulinum neurotoxins (BoNTs) are some of the most potent protein toxins and their modular architecture have made them molecules of choice for the design of a novel class of biopharmaceuticals labelled Targeted Secretion Inhibitors. The intracellular activity of BoNT towards members of the soluble N-ethylmaleimide-sensitive factor-attachment protein receptor family, a critical element of vesicular secretion, can be redirected towards specific cell types. This concept relies on the functionality of the LHn fragment which consists of the catalytic and translocation domains of BoNT. The distinct but variable properties of each serotype, such as substrate specificity and duration of action, have led to the development of tailor-made TSI with optimised applicability. Recent report of the biological activity of SXN101959 in inhibiting GH secretion has demonstrated the significance of using LHn/D as a framework for protein engineering. The crystal structures of the LHn fragment and SXN101959 present for the first time the properties of the combined translocation and catalytic domains of BoNT/D. The data provided have highlighted the plasticity of the light chain and its extensive interaction with the translocation domain. The translocation domain appears as a 'solid scaffold' on which to build on a retargeting partner. Indeed the structure of SXN101959 showed that addition of a long peptide ligand in a central position did not cause any significant structural conformational changes, thus demonstrating the robustness of LHn structural framework, and its ability to accommodate novel arrangements of the domain subunits, such as central presentation, within TSI proteins. Finally, the solution scattering analysis confirmed the structural integrity of the LHn fragment across serotypes. Altogether this study has provided useful information on the structure of BoNT/D, and the LHn fragment in general, that will be essential for the development of the next generation of protein therapeutics based on the TSI technology.

Methods

LHn/A, /B and /D cloning, expression and purification. Cloning, expression and purification were carried out following methods described previously[17,18]. Briefly, LHn constructs were cloned into modified pET26b expression vector (Novagen, Madison, WI, USA) with a C-terminal 6 x His-tag and expressed into *E. coli* BL21 (DE3) cells. The LHn genes were engineered to encode for an enterokinase or factor Xa cleavage site between the LC and Hn domains. Proteins were purified by affinity chromatography using Ni^{2+}-charged chelating sepharose column (GE Healthcare, UK) and hydrophobic interaction (Phenyl sepharose column, GE Healthcare, UK). Samples were kept in 0.05M HEPES pH 7.2 with 0.2 M NaCl.

The LHn/D construct used for X-ray crystallography correspond to the catalytically inactive LHn/D double-mutant E230Q and H233Y. Standard site directed mutagenesis techniques employing PCR amplification and subsequent sequencing were used to create the mutated DNA construct. In addition to the mutations required to eliminate catalytic activity, this construct contains an engineered enterokinase cleavage site between LC and Hn but was kept as a single chain sample for crystallisation. Expression and purification were performed as per the other LHn samples described above.

SXN101959 preparation. This protein was prepared as previously reported[15]. SXN101959 was designed using the LC and Hn of BoNT/D (Entrez protein database accession number P19321), a GHRH ligand domain, a lead sequence spacer, and a factor Xa protease activation site located between the two domains, as previously reported.

GHRH receptor activation assay (cAMP second messenger accumulation). Frozen CHO-K1 cells recombinantly expressing the human GHRH receptor were rapidly thawed, collected by centrifugation and re-suspended in assay buffer Hams F-12 (Life Technologies) , 5 mM HEPES (Life Technologies), 1 mM 3-isobutyl-1-methylxanthine (IBMX, from Calbiochem), 0.05% BSA (Perkin Elmer), pH 7.4). Cell suspension (5,000 cells well-1) was incubated in white 384-well microplates in triplicate with increasing concentrations of human GHRH(1–44, from Bachem) (final concentration 1 pM—100 nM) or SXN101959 (10 pM—1 μM), which had been serially diluted in buffer. Plates were incubated at 21 ± 3 °C for 90 minutes. Following ligand incubation, detection mix (consisting of cell lysis buffer, Lance Eu-W8044–labelled streptavidin, cAMP-biotin, and Alexa Fluor 647–anti-cAMP antibody) (kit from Perkin Elmer) was added to each well. Plates were incubated at room temperature for 24 hours, and the fluorescence emitted at 665 and 615 nm following excitation at 320 nm was measured using an EnVision microplate reader (PerkinElmer). Raw data was plotted as 615/665 nm emission ratio and converted to percentage of the maximum response (Emax) defined by hGHRH(1–44). Dose response curves were fitted using a variable slope model (using GraphPad Prism software) using equation (1).

$$Y = Bottom + \frac{Top - Bottom}{1 + 10^{(LogEC50 - X)n_H}} \quad (1)$$

where Y is the RFU, top and bottom are plateaus in the units of the Y axis, X is the log molar concentration and n_H is the Hill slope. EC50 value is the negative log of the mid-point location of the curve.

X-ray crystallography. Crystals of LHn/D were grown with 2 μl of protein at 56 mg/ml mixed with 1 μl of reservoir solution consisting of 30% v/v pentaerythritol ethoxylate (15/4 EO/OH), 0.1 M HEPES 7.5, 6% w/v polyvinylpyrrolidone (MIDAS D2, Molecular Dimensions) and suspended above the well as a hanging drop. Crystals grew within 7–10 days into pear-like shapes of 50 to 500 μm in length.

SXN101959 crystals were grown from the sitting drop method against 0.1 M sodium actetate pH 5.5, 0.8 M sodium formate, 10% w/v polyethylene glycol 8000, 10% w/v polyethylene glycol 1000 (Clear Strategy Screen 1, B6, Molecular Dimensions). A drop of 200 nl of protein at 10 mg/ml was mixed with equal amount of reservoir and incubated at 16 °C.

Diffraction data were collected at station I03 of the Diamond Light Source (Oxon, UK), equipped with a PILATUS-6M detector (Dectris, Switzerland). A complete dataset to 2.3 Å and 3.1 Å were collected from single crystals at 100 K for LHn/D and SXN101959, respectively. Raw data images were processed and scaled with DIALS[32], and AIMLESS[33] using the CCP4 suite 6.5[34]. The resolution cut-off chosen was based on the $CC_{1/2}$[35] as adding weak high-resolution data beyond the commonly used limits recorded on the highly sensitive detector showed significant improvements in electron density maps. For LHn/D, initial molecular replacement was performed with the coordinates of LC/D (PDB code 2FPW[23]) to determine initial phases for structure solution in PHASER[36]. The translocation domain was built in from a partial poly-alanine model of the secondary structures of LHn/B (PDB code 2XHL[18]) and extended with BUCCANEER[37]. The refined LHn/D structure was then used for molecular replacement to solve SXN101959. The working models were refined using REFMAC5[38] and manually adjusted with COOT[39]. Water molecules were added at positions where Fo−Fc electron density peaks exceeded 3σ, and potential hydrogen bonds could be made. Validation was performed with MOLPROBITY[40]. Crystallographic data statistics are summarized in Table 1. All figures were drawn with PyMOL (Schrödinger, LLC, New York).

Small-angle X-ray scattering. Protein samples of LHn/A, LHn/B, and LHn/D were kept in 0.05 M HEPES 7.2, 0.2 M NaCl for SAXS analysis and checked for monodispersity by dynamic light scattering (Malvern Instruments). Samples were diluted to give three final concentrations at 5, 2.5 and 1 mg/ml. These dilutions were checked by absorbance measurement at 280 nm using a NanoDrop ND-1000 (Thermo Scientific), and concentration discrepancies were corrected for data analysis.

Solution scattering (SAXS) data for the three LHn molecules were collected at DESY, (Germany), EMBL Hamburg, beamline X33[41]. For each sample, a buffer measurement was performed before and after the sample at room temperature, using a PILATUS-1M (Dectris, Switzerland) detector. SAXS measurements were performed for each sample at three concentrations. Calibration was carried out with BSA at 5 mg/ml in 0.05 M HEPES pH 7.5 as a molecular weight standard. The data were processed and

averaged with PRIMUS[42] followed by particle distance distribution function *P(r)* analysis using GNOM[43]. Data were further treated by using DAMMIF[44] to produce twenty *ab initio* models that were averaged with DAMAVER[45].

References

1. Schiavo, G., Matteoli, M. & Montecucco, C. Neurotoxins affecting neuroexocytosis. *Physiol. Rev.* **80,** 717–766 (2000).
2. Chen, S. Clinical uses of botulinum neurotoxins: current indications, limitations and future developments. *Toxins (Basel)* 4, 913–939 (2012).
3. Montal, M. Botulinum neurotoxin: a marvel of protein design. *Annu. Rev. Biochem.* **79,** 591–617 (2010).
4. Südhof, T. C. & Rothman, J. E. Membrane fusion: grappling with SNARE and SM proteins. *Science* **323,** 474–477 (2009).
5. Steinman, A., Galon, N., Arazi, A., Bar-Giora, Y. & Shpigel, N. Y. Cattle immune response to botulinum type D toxoid: results of a vaccination study. *Vaccine* **25,** 7636–40 (2007).
6. Coffield, J. A. *et al. In vitro* characterization of botulinum toxin types A, C and D action on human tissues: combined electrophysiologic, pharmacologic and molecular biologic approaches. *J. Pharmacol. Exp. Ther.* **280,** 1489–1498 (1997).
7. Eleopra, R. *et al.* Botulinum neurotoxin serotype D is poorly effective in humans: an *in vivo* electrophysiological study. *Clin. Neurophysiol.* **124,** 999–1004 (2013).
8. Kroken, A. R., Karalewitz, A. P.-A., Fu, Z., Kim, J.-J. P. & Barbieri, J. T. Novel ganglioside- mediated entry of botulinum neurotoxin serotype D into neurons. *J. Biol. Chem.* **286,** 26828–26837 (2011).
9. Peng, L., Tepp, W. H., Johnson, E. A. & Dong, M. Botulinum neurotoxin D uses synaptic vesicle protein SV2 and gangliosides as receptors. *PLoS Pathog.* **7,** e1002008 (2011).
10. Schiavo, G. *et al.* Identification of the nerve terminal targets of botulinum neurotoxin serotypes A, D, and E. *J. Biol. Chem.* **268,** 23784–23787 (1993).
11. Naumann, M., Boo, L. M., Ackerman, A. H. & Gallagher, C. J. Immunogenicity of botulinum toxins. *J. Neural. Transm.* **120,** 275–290 (2013).
12. Masuyer, G., Chaddock, J. A., Foster, K. A. & Acharya, K. R. Engineered botulinum neurotoxins as new therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **54,** 27–51 (2014).
13. Somm, E. *et al.* A botulinum toxin-derived targeted secretion inhibitor downregulates the GH/IGF1 axis. *J. Clin. Invest.* **122,** 3295–3306 (2012).
14. Leggett, J. *et al.* GHRH receptor-targeted botulinum neurotoxin selectively inhibits pulsatile GH secretion in male rats. *Endocrinology* **154,** 3305–3318 (2013).
15. Somm, E. *et al.* Comparative inhibition of the GH/IGF-I axis obtained with either the targeted secretion inhibitor SXN101959 or the somatostatin analog octreotide in growing male rats. *Endocrinology* **154,** 4237–4248 (2013).
16. Foster, K. A. *et al.* Re-engineering the target specificity of Clostridial neurotoxins - a route to novel therapeutics. *Neurotox. Res.* **9,** 101–7 (2006).
17. Masuyer, G. *et al.* Crystal structure of a catalytically active, non-toxic endopeptidase derivative of *Clostridium botulinum* toxin A. *Biochem. Biophys. Res. Commun.* **381,** 50–53 (2009).
18. Masuyer, G., Beard, M., Cadd, V. A., Chaddock, J. A. & Acharya, K. R. Structure and activity of a functional derivative of *Clostridium botulinum* neurotoxin B. *J. Struct. Biol.* **174,** 52–57 (2011).
19. Lacy, D. B., Tepp, W., Cohen, A. C., DasGupta, B. R. & Stevens, R. C. Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nat. Struct. Biol.* **5,** 898–902 (1998).
20. Swaminathan, S. & Eswaramoorthy, S. Structural analysis of the catalytic and binding sites of *Clostridium botulinum* neurotoxin B. *Nat. Struct. Biol.* **7,** 693–699 (2000).
21. Kumaran, D. *et al.* Domain organization in *Clostridium botulinum* neurotoxin type E is unique: its implication in faster translocation. *J. Mol. Biol.* **386,** 233–245 (2009).
22. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797 (2007).
23. Arndt, J. W., Chai, Q., Christian, T. & Stevens, R. C. Structure of botulinum neurotoxin type D light chain at 1.65Å resolution: repercussions for VAMP-2 substrate specificity. *Biochemistry* **45,** 3255–3262 (2006).
24. Guo, J. & Chen, S. Unique substrate recognition mechanism of the botulinum neurotoxin D light chain. *J. Biol. Chem.* **288,** 27881–27887 (2013).
25. Agarwal, R., Schmidt, J. J., Stafford, R. G. & Swaminathan, S. Mode of VAMP substrate recognition and inhibition of *Clostridium botulinum* neurotoxin F. *Nat. Struct. Mol. Biol.* **16,** 789–794 (2009).
26. Breidenbach, M. A. & Brunger, A. T. Substrate recognition strategy for botulinum neurotoxin serotype A. *Nature* **432,** 925–929 (2004).
27. Brünger, A. T. *et al.* Botulinum neurotoxin heavy chain belt as an intramolecular chaperone for the light chain. *PLoS Pathog.* **3,** 1191–1194 (2007).
28. Sikorra, S., Henke, T., Galli, T. & Binz, T. Substrate recognition mechanism of VAMP/synaptobrevin- cleaving clostridial neurotoxins. *J. Biol. Chem.* **283,** 21145–21152 (2008).
29. Galloux, M. *et al.* Membrane Interaction of botulinum neurotoxin A translocation (T) domain. The belt region is a regulatory loop for membrane interaction. *J. Biol. Chem.* **283,** 27668–27676 (2008).
30. Barabutis, N. & Schally, A. V. Growth hormone-releasing hormone: extrapituitary effects in physiology and pathology. *Cell Cycle* **9,** 4110–4116 (2010).
31. Svergun, D., Barberato, C. & Koch, M. H. J. Crysol - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28,** 768–773 (1995).
32. Gildea, R. J. *et al.* New methods for indexing multi-lattice diffraction data. *Acta Crystallogr.* **D70,** 2652–2666 (2014).
33. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr*. **D62,** 72–82 (2006).
34. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta. Crystallogr.* **D50,** 760–3 (1994).
35. Evans, P. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr.* **D69,** 1204–1214 (2013).
36. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
37. Cowtan, K. Fitting molecular fragments into electron density. *Acta Crystallogr*. **D64,** 83–89 (2008).
38. Murshudov, G. N. *et al.* Refmac5 for the refinement of macromolecular crystal structures. *Acta Crystallogr.* **D67,** 355–67 (2011).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr.* **D66,** 486–501 (2010).
40. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr.* **D66,** 12–21 (2010).
41. Blanchet, C. E. *et al.* Instrumental setup for high-throughput small- and wide-angle solution scattering at the X33 beamline of EMBL Hamburg. *J. Appl. Cryst.* **45,** 489–495 (2012).

42. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36,** 1277–1282 (2003).
43. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **25,** 495–503 (1992).
44. Franke, D. & Svergun, D. I. Dammif, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.* **42,** 342–346 (2009).
45. Volkov, V. V. & Svergun, D. I. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Cryst.* **36,** 860–864 (2003).

Acknowledgements
We thank the scientists at station IO3 of Diamond Light Source, Didcot, Oxfordshire (UK) for their support during X-ray diffraction data collection and station X33 at DESY, (Germany), EMBL Hamburg during SAXS experiments. K.R.A. wishes to thank Ipsen Bioinnovation Limited for a Research Fellowship and J.R.D. is supported by a joint post-graduate studentship between University of Bath and Ipsen Bioinnovation Limited. We are grateful to Dr. Mark Elliott and Dr. Keith Foster (both from Ipsen Bioinnovation Limited) for providing the receptor activation data for SXN101959 and for the constructive comments on the manuscript respectively.

Author Contributions
G.M. performed all protein expression, purification (except SXN101959), crystallography experiments, analysed the data and wrote the manuscript. J.R.D. carried out detailed structural analysis of SXN101959 and wrote the manuscript. K.M. expressed and purified SXN101959 protein used in this study. J.C. analysed the data and edited the manuscript. K.R.A. supervised the structural study, analysed the data and edited the manuscript. All authors reviewed the manuscript.

Additional Information
The atomic coordinates and structure factors (codes 5BQN and 5BQM) have been deposited in the Protein Data Bank (http://wwpdb.org).

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors G.M., J.R.D. and K.R.A. from the University of Bath declare no competing financial interests. J.C. and K.M are employees of Ipsen Bioinnovation Limited.

How to cite this article: Masuyer, G. *et al.* Structural analysis of *Clostridium botulinum* neurotoxin type D as a platform for the development of targeted secretion inhibitors. *Sci. Rep.* **5**, 13397; doi: 10.1038/srep13397 (2015).